THE PHYSICAL SOCIETY

REPORTS ON PROGRESS IN PHYSICS

VOLUME IX (1942–43)

GENERAL EDITOR
W. B. MANN

Price to non-Fellows 25s. net

Reproduced and Printed Photo-Litho Offset for
WM. DAWSON & SONS LTD.
16, West Street, Farnham, Surrey
1960

With the permission of
THE PHYSICAL SOCIETY

Dutch reprint

CONTENTS

PAGE

Preface iv

Reports of the Gassiot Committee of the Royal Society:

Atmospheric physics and chemistry. By SYDNEY CHAPMAN . . . 1

The spectrum of the Sun in the far ultra-violet. By A. HUNTER . . 5

Absorption spectra and absorption coefficients of atmospheric gases. By W. C. PRICE 10

The infra-red spectra of atmospheric gases other than water vapour. By G. B. B. M. SUTHERLAND and G. S. CALLENDAR 18

The absorption of water vapour in the far infra-red. By T. G. COWLING 29

The spectrum of the night sky. By R. W. B. PEARSE 42

Ionization phenomena in the earth's atmosphere. By J. SAYERS . . 52

The properties of neutral and ionized atomic oxygen and their influence on the upper atmosphere. By H. S. W. MASSEY and D. R. BATES . 62

Photochemical processes in an oxygen-nitrogen atmosphere. By C. H. BAMFORD 75

The photochemistry of atmospheric oxygen. By SYDNEY CHAPMAN . 92

Solar physics. By A. HUNTER 101

The structure and elasticity of rubber. By L. R. G. TRELOAR 113

Recent research on counter tubes. By J. D. CRAGGS 137

Insoluble monolayers at the air-water and oil-water interfaces. By A. E. ALEXANDER 158

Field emission of electrons. By R. O. JENKINS 177

Physics and the search for oil. By J. McG. BRUCKSHAW 198

Sound. By E. G. RICHARDSON 228

Experimental study of x-ray scattering in relation to crystal dynamics. By KATHLEEN LONSDALE 256

Theoretical investigations on the relation between crystal dynamics and x-ray scattering. By MAX BORN 294

The physics of paintings. By F. IAN G. RAWLINS 334

Author Index 349

PREFACE

In order to conform to the need for economy of time and material, the present volume of these *Reports* has been extended in scope to cover the two years 1942 and 1943. This policy of publishing one volume every second year instead of annually has been adopted for the duration of the war, but it is to be hoped that even so Volume X will appear before the autumn of 1945 and that annual publication will by then have been resumed.

In view of certain misunderstandings which have been brought to my notice, a few words of explanation should perhaps be given regarding the amended form of the Harvard system now being used in these *Reports*. This system, as it is generally applied, is most eminently suited to ordinary scientific publications containing but few references in all. In the case of a review article, however, this is not necessarily so, for there are many more references to the original publications, and the text tends to become cluttered with a miscellany of names and dates. To reduce this disadvantage to a minimum an abbreviated form of the Harvard system has therefore been adopted, and "Brown (1941 a–43)" refers to *all* the papers published by Brown between the 1941 a and 1943 references. Reference to a few specific papers can be made simply as "Brown (1941 a, 41 c, 42)" and so forth.

Instead of our usual pedagogic *Report*, one has been kindly contributed by the Scientific Adviser to the Trustees of the National Gallery on *The Physics of Paintings*. A corrigendum slip for insertion between pages 134 and 135 of Volume VIII is also being circulated with this volume.

The production of Volume IX has been a race against time; responsibility for any shortcomings as to conformity and consistency I must apologetically accept and attribute to other diversions. Its appearance at this time may be accepted as a token of the will to "Progress in Physics" which may, I hope, find less restricted expression in the next volume of this series.

In conclusion, I wish to express my thanks to Dr. R. W. B. Stephens for kindly assisting in the editing of one of the *Reports*, and also to my fellow officers, J. H. Awbery and W. Jevons, for friendly help in many and various ways.

W. B. M.

Imperial College,
London, S.W. 7.
September 1943.

AUTHOR INDEX

Abbott, C. G., 19.
Abbott, F. R., 190.
Aborg, 173.
Adam, 158, 160, 161, 162, 164.
Adel, 10, 12, 19, 23, 24, 25, 26, 27, 37, 39, 40.
Adler, 199, 202, 217, 223.
Ahearn, 187, 194.
Akamatu, 169.
Alden, 259.
Aldrich, 37.
Aleksandrov, 125.
Alexander, 161–164, 167, 170–174.
Alfrey, 121, 124.
Alfvén, 107.
Allan, 249, 250.
Allison, 258, 260, 281.
Allsopp, 14.
Amemiya, 66.
Anderson, 53.
Andrade, 240.
Anon., 337, 338, 340.
Antevs, 225.
Appleton, 54, 56, 57, 60, 61, 72, 73.
Arnold, 253.
Arnott, 239.
Askew, 172.
Astbury, 165.
Avery, 26.
Aybar, 234.

Babcock, H. D., 43, 75.
Babcock, H. W., 44, 105.
Bacher, 62.
Badger, 75.
Badgley, 190.
von Bahr, 34, 35, 37, 39.
Bailey, 26.
Bakker, 148.
Baltzer, 258, 278, 280, 287, 296.
Bamford, 80.
Bär, 235.
Baranek, 245.
Baranev, 166.
Barker, 19, 23, 26, 37.
Barnes, N. F., 344.
Barnes, S., 152.
Bartholemé, 26.
Bartholomew, 247.
Bartlett, 63.
Barton, 207, 223.
Bate, 230, 243, 244.
Bateman, 209.
Bates, 65–67, 69, 72, 88.
Baxter, 81, 257.
Bay, 153.
Beams, 188.
Bear, 88.
Beck, 109.
Becker, 192–196.
Beckett, 337.
Beer, 215.
Begbie, 308, 327.
Begun, 252.
Beischer, 168.
Bekkedahl, 119, 121, 123, 124.
Bell, 187.
Bender, 233.
Benedict, 13.
Benfield, 218.
Benjamin, 192–196.
Bennett, 187.
Beretta, 77, 78.
Berkner, 57.
Bernal, 29.
Bernard, 43, 44.
Bernstein, 167.
Berry, 257.
Bertsch, 267.
Best, 57.
Beutler, 89.
Bhagavantam, 283.
Bhimasenachar, 235, 283.
van Biesbroeck, 42.
Bigelow, 340.
Bikerman, 167.
Bilhuber, 238.
Bilinsky, 263.
Bishui, 282.
Blackman, 294, 304, 309.
Blake, 263.
Blanc, 257.
Bleuler, 302, 317.
Bloch, 258.
Blodgett, 165, 167.
Boas, 281.
Bodenstein, 85, 88.
Bolt, 245, 247.
Boltzmann, 128.
Boner, 243.
Bonhoeffer, 77, 87, 88.
Bonner, 12.
Bopp, 230.
Born, 199, 264, 268, 277, 282, 285, 294–296, 299, 301, 303, 304, 307–309, 312, 314, 318–322, 327, 328.
Borrmann, 257.
Bosworth, 195.
Bowden, 169.
Bowen, 43.
Bowie, 223.
Boyce, 12, 14, 63.
Boyd, 162–164, 170.
Bradbury, 52.
Bragg, 294, 319, 321.
Bramley, 151.
Brentano, 257.
Bresler, A. E., 160.
Bresler, S. E., 159, 160.
Brewer, 151.
Bridgman, 281, 286.
Brill, 266.
Brillouin, 250, 295.
Brindley, 260, 261, 264, 265, 274.
Briscoe, 160.
Brode, 15.
Brooks, 36.
Brown, R. E., 160, 170.
Brown, R. L., 245, 247.
Broz, 267.
Bruce, 207, 209, 287, 288.
Bruckshaw, 218.
Bryan, 218.
Buchwald, 108.
Buckingham, 65, 67.
Burmeister, 26.
Burris-Meyer, 248.
Burton, 10, 14.

Cabannes, 48, 49.
Callendar, 10, 21, 23.
Campbell, J. S., 63.
Campbell, R. W., 87.
Carlisle, 81.
Cary, 163.
Chadwick, 151.
Chalonge, 11.
Chambers, 187.
Chapman, 10, 14, 56, 72, 73, 77, 93, 94, 97, 98, 100.
Choi, 26.
Choong, 11.
Christenson, 151.
Cillié, 7.
Clapp, 251.
Clapson, 132, 133.
Clark, 342.
Claus, 269, 288.
Clayton, 172.
Clements, 25.
Coehn, 86.
Colange, 11, 78.
Colby, 267.
Collins, 16.
Colwell, 239.
Compton, 178, 258, 260, 268, 287.
Condon, 63, 64.
Cook, 88.
Cooley, 20.
Coolidge, 178.
Copeland, 163, 164.
Corbiére, 81.
Cornell, 32, 35, 38.
Cosyns, 151.
Coulson, 259.
Coumoulos, 167.
Cowling, 107.
Craggs, 148, 150, 151.
Cravath, 52.
Crawford, 26.
Cremer, 247.
Crisp, 174.
Cross, 339, 340.
Curran, 143, 145.
Curry, [illegible].

Dahlstrom, 190.
Danforth, 151.
Daniel, 195.
Danielli, 172.
Daniels, 25, 81, 82, 85, 86.
van Dantzig, 337.
Daraiswami, 235.
Darwin, 294.
Davey, 154.
Davies, 187.
Davis, A. H., 251.
Davis, C. W., 224.
Dawsey, 11, 24, 82, 87.
Dawton, 266.
Dayal, 283.
De Bruyne, 181, 183, 187.
Debye, 259, 287, 294–296, 299, 302, 315, 317.
Déjardin, 43, 45, 73.
Del Rosario, 182, 187.
Deming, 62, 93–95, 97–100, 253, 267, 288.
Dennison, 26, 30, 31, 38.
Dervichian, 159, 160, 162–164.
Dickinson, 81.
Dieke, 25, 44, 75.
Dimmick, 248.
Dix, 210.
Dixon, 13, 25.
Doss, 169.
Drew, 248.
von Droste, 150.
D.S.I.R., 346.
DuBridge, 152.
Duell, 344.
Duffendack, 138, 152, 153.
Dufton, 249.
DuMond, 258, 259.
Duncan, 12, 14, 79, 80, 82, 84, 85.
Duncanson, 259.
Dunning, 143.
Dunworth, 150, 151.
Dupérier, 108.
Durand, 286.
Dushman, 178.
Dutheil, J., 11, 78.
Dutheil, M., 11.
Dutta, 24, 82.

Eagleson, 228, 229.
Ebers, 26.
Eckert, 35.
Eddington, 111, 112.
Edlén, 7, 101–104, 106.
Ehrhardt, 265.
Eisenstein, 153, 154.
Elbe, 87.
Elkins, 224.
Ellis, 151.
Ellison, 109, 110.
Elsasser, 33, 36, 38–40.
Elvey, 42–45, 48, 49, 73.
Epstein, 229, 237.
Eucken, 234.
Evans, E. J., 240.
Evans, R. D., 138, 151.
Everitt, 251.
Ewald, 311.
Ewing, 207–209.
Eyring, C. F., 189, 181, 185.
Eyring, H., 88, 163.

Fairbanks, 247.
Falckenberg, 34, 30.
Farmer, 57.
Farnsworth, 341.
Faxén, 295.
Fay, 243.
Feather, 151.
Ferri, 116, 119, 132.
Fessenkov, 105.
Field, 118, 120.
Finbak, 266.
Finkelnburg, 75.
Firestone, 251.
Fisher, 223.
Fleming, 190, 252.
Fletcher, 345.
Flory, 13, 80, 81, 83, 85.
Focke, 236, 237.
Forbes, 78, 79.
Ford, 170.
Forsaith, 338.
Försterling, 280.
Foster, 264.
Fourt, 122, 100, 103.
Fowle, 35–37, 39, 40.
Fowler, 29, 183–180.
Fox, F. E., 132, 236.
Fox, G. W., 267.
Fox, J. J., 20, 282.
Frank, 76.
Fraser, 267.
Frederick, 267.
Frerichs, 63.
Fricke, 234.
Friedman, 151.
Frongia, 42.

Gallagher, C., 239.
Gallagher, D. M., 168.
Galloway, 26.
Ganitta, 243.
Gardiner, 248.
Gauzit, 42, 45, 48–50.
Gaviola, 89.
Gaydon, 45.
Gee, 125.
von Geel, 147.
Gehman, 120.
Geiger, 143.
Gerhard, 23.
Germer, 167.
Gettens, 339, 340, 344, 349.
Gibbs, 158.
von Gillhaussen, 168.
Gilliland, 57, 60.
Gingrich, 153, 154, 280, 281.
Ginsbury, 31, 38.
Gish, 53.
Glazov, 166.
Goens, 264, 286.
Goetz, 266.
Goldstone, 190.
Goodeve, 14.
Göppert-Mayer, 204, 308.
Gorodetzki, 35.
Gossling, 178, 180, 181, 186, 187.
Götz, 92.
Goudsmit, 62.
Gramath, 86.
Greaves, 6.
Green, 211.
Gregg, E. C., 235.
Gregg, R. Q., 280, 281.
Gregg, S. J., 168.
Greiner, 144.
Greisen, 151.
Griffith, R. L., 266.
Griffith, T. R., 114.
Groom, 345.
Groth, 77, 83, 86.
Grotrian, 102.
Grüneisen, 264.
Gupta, 282.
Gurney, 139.
Gutenberg, 206, 223.
Guth, 114, 126, 132.

Haalch, 218.
Haber, 87.
Haefer, 195.
Hagstrum, 64, 65, 71.
Haines, 138.
Hall, J. B., 26.
Hall, J. H., 281.
Hammer, 223.
Harding, 154.
Hardy, 26.
Harkins, 159, 160, 162–164.
Harnwell, 59.
Harper, 149.
Harris, L., 13, 81.
Harris, M., 122.
Harris, P. M., 266.
Harris, S., 267.
Harteck, 77, 78, 87.
Hartley, 218.
Hartmann, G. K., 230.
Hartmann, J., 241.
Hartree, B., 64, 266.
Hartree, D. R., 64.
Harvey, 260, 264, 288, 289, 296.
Haskell, 214.
Haskins, 168, 169.
Hassel, 266.
Hayford, 223.
Heidt, 78, 79.
Helmholtz, 245.
Henderson, 190.
Hendricks, 266.
Henning, 11.
Henri, 14, 81.
Henrique, 80, 82, 84, 85.
Henry, 79.
Herget, 232.
Hermann, 266.
Herzberg, 13, 21, 22, 75, 76, 101.
Herzfeld, 230.
Hettner, 23, 24, 35, 37–39.
Heymann, 166.
Hibben, 86.
Hicks, 258, 259.
Higgs, 8.
Hilton, 239.
Hinshelwood, 84.
von Hippell, 144.
Hoffman, 170.
Holley, 167.
Hollnagel, 40.
Holloway, 139.
Holmes, 25, 81, 82.
Holweck, 218.
Honeyburne, 257.

Hönl, 260.
Hopfield, 16.
Hottel, 36.
Houston, C. E., 209.
Houston, W. V., 183.
Howard, 264, 265.
Howell, 199.
Hudspeth, 146.
Huggins, 135.
Hughes, 266.
Hull, 178.
Hunt, 245, 252.
Hunter, A., 16, 110.
Hunter, J. L., 236.
Huppertsberg, 153.

Inui, 66.
Iona, 264, 327.
Irons, 244.
Isemura, 169.
Ising, 218.
van Itterbleek, 234.

Jack, 247.
Jackson, 87.
Jacobs, 266.
Jahn, 285, 286, 296, 311, 312, 314, 315, 325.
Jakovleva, 11, 78.
James, H. M., 132, 205.
James, R. W., 174, 179, 261.
Janin, 44.
Jaques, 207.
Jauncey, 257, 258, 260, 267–269, 280, 283, 287, 288, 296, 321, 322.
Jenkins, F. A., 10.
Jenkins, R. O., 192–196.
Johnson, A. E., 224, 225.
Johnson, C. A., 238.
Johnson, C. H., 204.
Johnson, R. C., 44.
Johnson, R. P., 191.
Johnson, V. A., 265.
Johnston, E. H., 86.
Johnston, H. L., 13, 80, 81, 83, 85.
Joly, 159.
Jones, A. T., 244.
Jones, E. J., 14, 82.
Jordan, E. C., 251.
Jordan, V. L., 245.
Judson, 60.

Kaiser, 151.
Kakiuchi, 267.
Kanayama, 253.
Kanne, 142.
Kanta, 145.
Kaplan, 43, 45, 48.
von Kármán, 295, 304, 309.
Kastner, 243.
Kawashima, 247, 251.
Kaye, 246.
Kean, 199.
Keller, 234.
Kellermann, 264, 304, 309, 327.
Kellogg, 248.
Kendall, 233.
Kenty, 69.
Kerkum, 147.
Kern, 295.
Kernbaum, 86.
Kerr, 253.
Kiepenheuer, 5.
Kimball, 88.
King, 13.
Kirby, 57, 60.
Kirchhoff, 245.
Kirkpatrick, H. A., 258, 259.
Kirkpatrick, P., 258, 262, 280, 281.
Kirkwood, 159.
Kistiakowsky, 25, 77, 78, 87, 88.
Klauer, 267.
Klipsch, 204, 252.
Knaggs, 283, 284.
Knauss, 243.
Kneser, 284.
Knott, C. G., 199.
Knott, G., 167.
Knowler, 250.
Knudsen, 234, 249.
Kobayashi, 247.
Koenigsberger, 224.
Kohler, 264.
Kohn, 152.
Kolhörster, 151.
Kondratjew, 11, 78.
Kopsch, 78, 87.
Korff, 139–143, 145, 154.
Kornfield, 88.
Kossel, 257.
Kotani, 65.
Kracek, 266.
Kreuchen, 152.
Krishnan, 235.

Labaw, 244.
Ladenburg, 12, 14, 23, 33.
Lamb, 314.
Lambrey, 11, 13, 80, 81.
Lampland, 19, 24, 37, 39.
Lancaster-Jones, 108, 223.
Langevin, 230.
Langley, 19.
Langmuir, 160, 161, 165, 166, 178.
Lark-Horovitz, 265.
Lauchli, 11.
von Laué, 257, 268, 294–296, 317.
Laurie, 334, 342.
Lauritsen, 180, 181, 185.
Laval, 257, 267, 269, 286, 288, 289, 296, 318, 326.
Lavin, 87.
Lawlor, 215.
Lazarus, 241.
Lazurkin, 125.
Leben, 169.
Ledermann, 309.
Lees, 146.
Leet, 207–209.
LeGalley, 152, 153.
Leifson, 12, 13, 80, 81.
Leighton, 10, 14, 76.
Leithauser, 24.
Lejay, 218.
Lenihan, 240.
Leonard, 234.
Lester, 215, 216.
Levin, 26.
Levine, 223.
Lewis, B., 89.
Lewis, D., 247, 252.
Lewis, W. B., 138, 143, 149.
Lifschutz, 138.
Lifshitz, 248.
Lilienfeld, 177, 178, 180.
Lindahl, 250.
Lindsay, 244.
Linke, 42–45, 48, 49, 73.
Linnett, 26.
Lipson, 260.
Livingston, 139.
Lloyd, 249.
Locher, 149, 152–154.
Lochte-Holtgreven, 44.
Long, 169, 171.
Lonsdale, 264, 269, 270, 277, 283–285, 288, 294, 297, 310–313, 317–319, 321, 322, 324–326, 328.
Lottermoser, 249.
Lowry, 86.
Lozier, 64, 65, 71.
Lynton, 225.
Lyon, 341.

Maa, 250.
McBain, 169–171.
McDermott, 201.
Macdonald, 79, 80.
McGinnis, 239.
Mack, 114.
Mackeown, 181.
McNatt, 257, 258, 287–290.
McNish, 224, 225.
Macready, 225.
Makita, 253.
Malter, 148.
Mangelsdorf, 36.
Marager, 334.
Margenau, 31.
Mariens, 234.
Mark, 114, 118, 119, 121, 122, 124, 126.
Marro, 248.
Marsden, 161.
Martin, A. E., 282.
Martin, P. E., 23.
Martin, S. T., 196.
Martyn, 8, 54, 55.
Mason, 250.
Massey, 55, 56, 65–67, 69, 72.
Matheson, 123, 124.
Matumura, 253.
Matz, 242.
Maugin, 296.
May, 138, 143, 147, 269.
Mecke, 29, 75, 81, 93, 95.
Medicus, 146, 148.
Meirner, 149.
Meloin, 14, 78, 95.
Melville, 77.
Menzel, 7, 63, 105.
Meres, 199.
Meyer, C. F., 26.
Meyer, E., 5.
Meyer, Erwin, 240.
Meyer, K. H., 114, 115, 117, 119, 132.
Mezhueva, 165.
Miller, 243, 264, 280.
Millikan, 180–183, 185.

Millington, 88.
Mills, 170, 171.
Milne, 97, 106, 111.
Minkowski, 8.
Misra, 304.
Mizell, 122.
Mohler, 69, 71.
Moir, 250.
Montgomery, C. G., 140, 143–145.
Montgomery, D. D., 140, 143–148, 150, 151.
Moore, D. H., 188.
Moore, W. J., 163.
Morris, 152, 153.
Morse, 247.
Mortimer, 253.
Mott-Smith, 205, 209.
Mouguin, 160.
Mouroff, 8.
Mugele, 138, 151.
Müller, A., 124, 283.
Müller, E. W., 177, 188, 189, 191, 192, 194, 195.
Mulliken, 15, 75.
Murray, 172.
Musgrave, 84.
Muskat, 199, 209.
Myers, 159.

Naismith, 54, 72.
Narasimhaiya, 235.
Nath, 281, 294.
Nayar, 283.
Neary, 152.
Neher, 149.
Nemejcová, 267.
Nereson, 151.
Nespor, 239.
Nettleton, 216, 218, 222.
Neuimin, 88.
Newman, 243.
Newton, 108, 109.
Nicolet, 42.
Nielsen, A. H., 26.
Nielsen, H. H., 26, 31.
Nilakantan, 281–283, 294, 321, 325.
Nippoldt, 223.
Nishikawa, 267.
Noges, 10, 14, 76, 77, 80, 82, 84, 85, 89.
Nordheim, 183–185.
Norgaard, 218.
Norling, 150.
Norrish, 81, 82.
Northrup, 244.
Norton, 57.
Nutting, 164, 169, 171
Ny, 11.

Obata, 247, 252.
Oberst, 242.
Oka, 237.
Oldenberg, 80.
Olson, 248.
Orbán, 152.
Ostwald, 344.
Oswatitsch, 229, 230.
Ott, 296, 297, 299.
Overbuck, 233.

Panisset, 345.
Parkinson, 247.
Payne, 63.
Pearse, 45.
Pearson, 87, 88.
Pease, 346.
Pelser, 115, 131.
Peng, 304.
Penman, 233.
Pentz, 206, 223.
Pepper, 221.
Perry, 169.
Peters, 266.
Peterson, 213, 234.
Petrie, 105.
Petrzilka, 143.
Pforte, 183.
Pickering, 149.
Pielemeier, 233.
Piersol, 183.
Pirson, 202.
Pisharoty, 282, 283, 285.
Placzek, 296, 297.
Plenderleith, 340.
Pokhil, 174.
Popov, 88.
Porter, 159, 167.
Posnjak, 266.
Power, 304.
Preston, G. D., 281, 294, 321.
Preston, W. M., 8, 16, 106.
Price, W. C., 10–14, 16, 77, 86.
Pugh, 211.
Pulley, 54, 55, 71.

Quarles, 188.
Quimby, 286, 325.

Rabinowitch, 76, 89.
Race, 168.
Rai, 65.
Rajewsky, 152.
Ralston, 170.
Ramaiya, 236.
Raman, 235, 268, 281–283, 294, 309, 321.
Ramsey, 138, 139, 144–146, 151, 155.
Randall, 31, 32, 35, 37, 38, 39.
Ransome, 211.
Rao, 236.
Ratcliffe, 57, 58.
Rathenau, 11, 86.
Rawlins, 26, 336, 337, 341–346.
Rayleigh, 73.
Read, 150.
Redfearn, 251.
Regener, E., 5.
Regener, H., 5.
Reiche, 33.
Renninger, 257.
Reynolds, 168.
Rice, 11, 82, 87.
Richardson, E. G., 235, 240, 247.
Richardson, O. W., 183, 185.
Richardson, R. S., 8.
Richter, 230.
Rideal, 26, 159–161, 163, 166.
Ridley, 260, 264, 265.
Riebar, 205.

Ries, 160.
Riley, 260.
Ritchie, M., 79.
Ritchie, P. D., 342.
Roberts, 127.
Robertson, 26, 282.
Robinson, 244.
Rocard, 230.
Rock, G. D., 236.
Rock, S. M., 215.
Rodebush, 84–87.
Rohloff, 238.
Rollefson, 10, 14, 76.
Romberg, 211, 212.
Rosaire, 199, 202, 217.
Rose, 138–142, 145.
Ross, 257, 262.
Rother, 182.
Roxburgh, 77.
Rschevkin, 246.
Rubens, 23, 40.
Ruheman, H., 339.
Ruheman, M., 123.
Rutherford, E., 143, 151.
Rutherford, H. M., 210.

Sabine, 247, 250.
Saha, 7, 65.
Sakata, 150.
Sakisaka, 267.
Salvatori, 202.
Sameshima, 169.
Sarginson, 268, 282, 285, 296, 299, 312, 315, 319–322, 325.
Saunders, 237.
Savic, 240.
Saxton, 233.
Sayers, 53, 56, 73.
Schaal, 236.
Schaefer, C., 23, 230.
Schaefer, V. J., 160.
Schallamach, 123, 124.
Scharwächter, 259.
Schein, 5.
Schenk, 88.
Schierkolk, 26.
Schilling, 187.
Schmid, 281.
Schmidt, 35, 36.
Schnaidt, 33, 34, 36.
Schneider, 11.
Schou, 14.
Schottky, 178, 180, 183, 185.
Schrödinger, 295, 296.
Schulman, 159, 165, 167.
Schumacher, 14, 23, 24, 77–79, 83, 85, 86.
Schwarzschild, 112.
Sebba, 166.
Seddon, 86.
Semenoff, 77.
Sen Gupta, 24, 79, 82.
Shankland, 259.
Shannon, 253.
Shaw, 223.
Shear, 237.
Shenstone, 102.
Sheppard, 132 133.
Sherman, 53.
Shive, 155.
Shockley, 191.
Sibaiya, 235.

Siedl, 237.
Sieg, 228, 229.
Siegel, 280, 286, 325.
Siegert, 222.
Simon, 123.
Simpson, 11–14.
Sirkar, 282.
Skipher, 143.
Skylarenko, 166.
Slawsky, 138.
Slipher, 7, 19, 37, 45.
Slotnick, 207, 208.
Smee, 148, 151.
Smith, C. S., 267, 296.
Smith, F. R., 154.
Smith, H., 269, 282, 284, 285, 286, 288, 294, 311, 321, 326.
Smith, H. A. 88.
Smyth, 59.
Snow, 26.
Snyder, 116, 117.
Sommerfeld, 259.
Spealman, 84, 85.
Spiers, 264.
Spinks, 21, 22.
Sponer, 12, 14.
Sprenger, 14, 85, 86.
Ställberg, 164, 165.
Stebbins, 267.
Steele, 212.
Stefanescu, 26.
Stein, 14.
Steinberg, 248.
Steiner, 75, 83.
Stenhagen, 164, 165, 167.
Stern, 186.
Stevens, 53.
Stevenson, 63.
Stever, 141, 143, 146, 148, 149, 154, 155.
Stewart, 239.
Stoddart, 86.
Stoll, 5.
Storks, 167.
Stout, 339, 340, 346.
Street, 145, 151.
Strong, 33, 37–39, 138, 143, 149.
Strothers, 145.
Struve, 42.
Stueckelberg, 59.
Stuhlinger, 141.
Subrahmanian, 282, 283.
Süchster, 242.
Suga, 10.
Sullivan, 153.
Sumato, 267.
Sundberg, 218.
Suryaprakasam, 235.
von Susich, 114.
Sutherland, 10, 12, 26, 168.
Swann, 140, 154.
Swings, 42–45, 48–50, 73, 101.
Swirles, 64.

Takamine, 10.
Talmud, B. A., 159.
Talmud, D. L., 159.
Tanaka, 10.
Tate, 64, 65, 71.
Teague, 281.
Telfair, 233.
Teorell, 161, 172, 173.
Terossipjantz, 246.
Thackeray, 101, 110, 111.
Thiessen, 119, 168.
Thompson, J. W., 26.
Thompson, R. A., 199.
Thompson, R. R., 232.
Tian, 86.
Tolman, 86.
Trapejnikov, 163.
Treloar, 117–122, 128–130.
Trichel, 143.
Trost, 142, 143, 146.
Tucker, 228, 229.
Tuckett, 124.
Tukada, 54.
Türk, 249.
Tuthill, 247.
Tutte, 168.
Tüxen, 64.
Twyman, 14.
Tyndall, 228.
Tyte, 243.

Ubbelohde, 284.
Ueberreiter, 125.
Unwin, 65, 67, 231.
Urelius, 218.
Urey, 11, 24, 82, 87.

Vacquier, 226.
Valko, 114.
Vassy, 93.
Vaughan, 77.
Vencov, 60.
Venhateswaran, 283.
Venkateswarlu, 235.
Verhoek, 85.
Vermeulen, 239.
Voges, 257.
Van Voorhis, 12, 14.

Wadlund, 279.
Wahl, 86.
Waldemeier, 8, 105, 107–109.
Wall, 128–133, 135.
Waller, I., 260, 279, 295, 296, 307, 315, 317, 322.
Waller, M., 239.
Warburg, 24, 75–78, 83.
Warner, 233.
Wartenberg, 40.
Waters, 214.
Watson, N. A., 249.
Watson, W., 21, 22.
Watson, W. W., 31.
Weatherby, 199.
Weber, E., 151.
Weber, L. R., 31, 35, 37–39.
Wehnelt, 187.
Weigle, 257, 267, 296, 302, 311, 317, 325, 326.
Weisz, 138, 139, 142, 143.
Wells, 167.
Wende, 87.
Werner, 147, 152.
White, 86.
Widdowson, 168.
Widness, 214.
Wiegand, 116, 117.
Wilde, 341.
Wilkening, 142.
Wilkes, 57.
Willard, 236.
Williams, P. S., 260, 288.
Williams, S. E., 8.
Wilson, 243.
Wilson, A. J. C., 260.
Wittstadt, 119.
Wolf, 214.
Wollan, 264.
Woo, 38, 39, 287.
Wood, A. B., 216.
Wood, L. A., 119, 121, 123.
Wood, R. W., 178.
Woodward, 145, 151.
Worley, 10.
Worthington, 253.
Wright, N., 32.
Wright, S. J., 286.
Wu, 26.
Wulf, 14, 62, 76, 78, 82, 93–95, 97–100.
Wyckoff, 219.
Wyman, 167.

Yakovleva, 11, 78.
Yamanonclu, 65, 66.
Yamina, 159.
Yeager, 243.
Yosida, 253.
Young, 162, 164.

Zabor, 80.
Zachariasen, 268, 279, 280, 283. 296, 304, 313, 323, 325
Zahl, 168.
Zanstra, 6.
Zeise, 85.
Zener, 259, 263, 264..
Zisman, 159.
Zohradnicek, 239.

ATMOSPHERIC PHYSICS AND CHEMISTRY*

By SYDNEY CHAPMAN, F.R.S.,
Imperial College, London

In 1941 the Secretary of State for Air (The Rt. Hon. Sir Archibald Sinclair, M.P.) instituted the Meteorological Research Committee,† to advise and assist in meteorological investigation. At present the Committee is concerned chiefly with problems closely concerned with the war, so that most of its work must, for a time, remain secret; but at its first meeting, in December 1941, the Committee recommended that the co-operation of the Royal Society should be invited, in regard to long-term fundamental meteorological research. The Air Ministry sent such an invitation, to which the Council of the Royal Society agreed, and the immediate responsibility for the work was entrusted to the Gassiot Committee,‡ whose terms of reference were suitably extended. The field of work specially proposed by the Air Ministry to the Royal Society is that of radiation in the atmosphere, particularly with reference to the radiation equilibrium conditions in the stratosphere; at present it is uncertain why the temperature of the stratosphere is 50° c. lower over the equator than over the poles, and why it varies considerably with time.

The principal gases in the atmosphere, nitrogen and oxygen, are very transparent to radiation of most wave-lengths, though the absorption of ultraviolet light by oxygen is of great importance in the ionosphere and down to the lower stratosphere. In the troposphere and stratosphere the radiative conditions are governed largely by such minor constituents as water vapour, carbon dioxide, ozone and perhaps other polyatomic constituents. To understand the radiation equilibrium at different levels in the atmosphere, it is necessary to know (*a*) what gases are present which have important absorption bands; (*b*) their absorption coefficients for different wave-lengths, and whether the absorption and emission spectra are continuous or have a fine-line structure; (*c*) the densities of these gases at various heights; and (*d*) how such gases as ozone are formed and decomposed. Such knowledge is needed as a basis for calculation of the radiative

* This article is introductory to the following nine *Reports*, which are printed in this volume by arrangement with the Royal Society.

† Constituted as follows: D. Brunt; S. Chapman (Chairman); G. M. B. Dobson; G. I. Taylor; the Director of the Meteorological Office; the Director of Scientific Research of the Ministry of Aircraft Production; the Director of the Naval Meteorological Service; and representatives of the Air Staff and of civil aviation.

‡ This Committee was first constituted in 1871, to administer an endowment fund for Kew Observatory, provided by J. P. Gassiot (1797–1877), who made various electrical experiments, and also founded the Scientific Relief Fund of the Royal Society. Other funds for British meteorological and magnetic observatories were later entrusted to the Committee, which also makes recommendations regarding the work of these observatories. At present the Committee is constituted as follows: E. V. Appleton, D. Brunt, S. Chapman, G. M. B. Dobson (Chairman), H. G. Lyons, H. S. W. Massey, G. C. Simpson, G. I. Taylor, G. T. Walker, the Astronomer Royal, the Director of the Meteorological Office, the President of the Royal Astronomical Society, the Treasurer (T. R. Merton) and Secretary (A. C. G. Egerton) of the Royal Society.

equilibrium temperature at different levels, and of its rate of change when conditions vary.

This field of study is very wide, and can be cultivated successfully only by co-operation between specialists of many kinds. Consequently the Gassiot Committee enlisted the aid of several workers* outside its own membership, in constituting three sub-committees to study particular branches of the subject, namely: (A) the chemical constitution of the atmosphere, at different heights; (B) the photophysics and photochemistry of the atmosphere, including its visible and ultraviolet spectroscopy; and (C) the infra-red spectroscopy and the radiation balance.

Atmospheric conditions at all heights depend largely on the sun's radiation, most of which can traverse the whole atmosphere to the earth's surface, if not stopped by clouds; but much of the ultraviolet part of the radiation is intercepted by the ionosphere and the ozone layer. Much of this absorbed radiation is not directly measurable, and must be inferred partly from atmospheric and laboratory data, and partly from our knowledge of solar physics. The Gassiot Committee has taken no narrow view of its field of work, and seeks to learn the full history of the solar radiation entering and passing through the atmosphere, as well as that of the outgoing radiation reflected or emitted by the earth's surface, the clouds, and the atmosphere itself. Hence the Committee's investigations include the ionosphere, where an important part of the sun's radiation is abstracted.

Under war conditions it is naturally difficult to carry on new researches into these problems, and during the first year of its work the Committee has confined itself mainly to a review of the present state of knowledge of the subject, partly in order to see where future efforts are most needed. On the recommendation of the Committee, however, the Royal Society appointed E. Glückauf as Mackinnon Research Student for two years, to work in collaboration with the Committee; he has been actively engaged in developing methods for the determination of the water content of air at different levels, and especially in the stratosphere, where the amount, which must be very small, is as yet unknown. Apparatus suitable for transport in aeroplanes is being developed to determine the concentration of water and ozone in the stratosphere. These researches fall within the scope of the chemical sub-committee (A), which has not contributed to the following nine *Reports*; it has, however, prepared plans for the determination of other atmospheric constituents also, including methane, ammonia, and the oxides of nitrogen, using both chemical and spectroscopic methods.

These chemical questions are discussed to some extent in the *Reports* by W. C. Price and by G. B. B. M. Sutherland and G. S. Callendar, which form part of a series of five spectroscopic *Reports*. This series begins with a discussion by A. Hunter of the sun's ultraviolet spectrum, for wave-lengths less than 2900 A., as to which no reliable direct evidence exists, or is likely soon to exist. Many phenomena, both solar and terrestrial, suggest that the sun radiates far more energy in the ultraviolet than would a black body at 6000° K., but it seems likely

* These include, besides the authors of the reports that follow, H. Dingle, K. G. Emeléus, E. Glückauf, A. R. Meetham, R. G. W. Norrish and F. A. Paneth.

that the enhancement is mainly in the far ultraviolet, beyond about 1200 A., where the intensity of the continuous spectrum may correspond to temperatures up to 20000° K.; in addition, there is likely to be very important line emission, especially at 1215 A. (due to H) and 584 A. (due to He). This enhancement is important for the ionosphere, but seems to have little bearing on the ozone layer.

The next three *Reports* summarize our present knowledge of the absorption spectra of the atmosphere and of many of its known or possible constituents; water vapour is treated by itself (in the *Report* by T. G. Cowling) because of its exceptional importance and the complexity of its infra-red spectrum. Limits can be set to the amounts of various possible constituents, by examination of the atmospheric absorption spectrum; and there is good hope that by using greater dispersion in the infra-red these limits can be made still more definite. It seems already clear, however, from the *Report* by Sutherland and Callendar, that the absorption and emission of long-wave radiation in the lower atmosphere is governed almost entirely by H_2O, CO_2 and O_3; this accords with past beliefs, but there remained a possibility, now excluded, that some other polyatomic constituent, present in small amount like ozone, might, like ozone, have an important influence on the radiation balance of the lower stratosphere.

The fifth spectroscopic *Report*, by R. W. B. Pearse, deals with the radiation emitted by the atmosphere itself, observable at night; this gives evidence of the presence of atomic oxygen and sodium in the upper atmosphere, and also of metastable molecular nitrogen; but much of the spectrum, including one strong ultraviolet band, remains uninterpreted. The night-sky spectrum throws light on recombination processes continually in progress in the higher layers, and gives some indication of the temperature there; but it does not tell us what constituents control the temperature by radiating away that part of the energy which is absorbed from ultraviolet sunlight and degraded to heat energy. At high levels ozone must be insignificant, and the amounts of water vapour and CO_2 are quite unknown (their infra-red emissivity will not be destroyed even if they are dissociated to OH (+H) and CO (+O)); but other gases, such as N_2O, may have a stronger influence on the temperature at such levels.

The next two *Reports* deal with the ionosphere; J. Sayers discusses the recombination processes by which ions and electrons reunite to form neutral atoms and molecules, and H. S. W. Massey and D. R. Bates summarize and discuss the properties of neutral and ionized oxygen atoms, which play so great a role in the ionosphere. Very much still remains obscure concerning the processes that go on there, by day and by night. They form a special group of the complex photochemical phenomena of the atmosphere, namely, those due to the absorption of the most energetic, far ultraviolet, quanta of solar radiation.

The eighth *Report*, by C. H. Bamford, summarizes the photochemical knowledge acquired in the laboratory, bearing on the processes that solar radiation may produce in a mixture of oxygen and nitrogen, containing also hydrogen (in H_2O and other forms). The possibilities are very numerous, but many can doubtless be ruled out as unimportant in the atmosphere. The last *Report*, by myself, describes the attempts made to elucidate one of the simplest and

most important problems of atmospheric photochemistry, namely, that of ozone. These studies are likely to be forerunners of similar discussions of the changing equilibrium of other minor constituents, such as the oxides of nitrogen, methane, and possibly the dissociation products of H_2O and CO_2, at different levels in the atmosphere. It seems probable that it will prove worth while to consider such questions even where they can have little immediate bearing on the main problem of the radiation equilibrium of the lower atmosphere; where direct observation is so difficult as in the stratosphere, indirect clues are of great value, and such are likely to result as a by-product of attempts at a detailed systematic synthesis of our knowledge of the upper atmosphere.

The following nine *Reports* will, it is hoped, contribute towards this end, and not least by bringing atmospheric problems to the notice of specialists in many fields, whose work bears on these problems, and who may feel moved to help in these studies.

THE SPECTRUM OF THE SUN IN THE FAR ULTRA-VIOLET

By A. HUNTER,
Royal Observatory, Greenwich

§1. INTRODUCTION

Our knowledge of the photochemistry of the Earth's atmosphere cannot be complete until we have a detailed knowledge of the radiation reaching the Earth from the Sun. That our comprehension of the processes occurring in the upper atmosphere is so fragmentary must be attributed in part to our defective knowledge of the distribution of solar radiation in the extreme ultra-violet. This *Report* draws attention to the few directions in which progress has been made in extending our knowledge of this fundamental meteorological factor.

§2. DIRECT OBSERVATION BELOW 2900 A.

Ozone absorption prevents entirely the observation of the solar spectrum at wave-lengths between about 2900 A. and 2200 A. at sea-level. Below 2000 A. oxygen absorption acts similarly. Surface observations at 1·7 km., using a spectrograph fed through a monochromator, have reached 2863 A. only by using 3-hr. exposures. Unmanned balloon observations at 30 km. by E. and H. Regener (1934) reached 2875 A., using much shorter exposures. Little progress can be expected in this direction.

It has frequently been suggested that solar radiation might be detected at wave-lengths around 2100 A., where ozone absorption is falling off and oxygen absorption is not too intense. Meyer, Schein and Stoll (1934) have actually detected radiation in this region with a sensitive photon counter; but Kiepenheuer (1937) failed to repeat their observations. Work of this sort, if calibrated and used in conjunction with accurate atmospheric absorption coefficients in this region (Meyer, 1941), might indicate the strength of the solar continuum in this region, but only the order of magnitude could be expected.

Speaking generally, then, no reliable direct evidence exists or is likely to exist in the near future as to the energy distribution in the solar spectrum below 2900 A.

§3. INFERENCES FROM DIRECT OBSERVATION ABOVE 2900 A.

It might be thought that extrapolation from the visible region or the near ultra-violet could give some indication of the energy distribution below 2900 A. Presumably at photospheric levels the radiation is black-body radiation; and, so long as the solar atmosphere does not absorb selectively, the radiation reaching the Earth will be distributed according to Planck's law. Observation in the visible region shows that this is roughly so, apart from Fraunhofer lines. This

indicates that in the visible region the total absorption coefficient of the material of the solar atmosphere is roughly independent of wave-length. There is now, however, good reason (Greaves, 1939) for believing that this results from a chance balancing of the contributions to solar atmospheric opacity from H absorption on the one hand and metallic absorption or H^- absorption on the other. This balance is unlikely to prevail at short wave-lengths. It is certain that in the near ultra-violet at the Balmer series limit 3646 A., a large discontinuity occurs in the solar continuum due to photo-electric ionization of 2-quantum H atoms. If there is sufficient continuum in the far ultra-violet spectrum of the Sun to show it, there will undoubtedly be a similar discontinuity below 911 A. due to photo-electric ionization of ground-level H atoms. Continuous absorption will, in fact, be expected at the principal series heads in the spectra of all elements, and these are mostly in the far ultra-violet. Quite apart from line absorption, then, we shall need continuous absorption coefficients in the far ultra-violet for all the constituents of the solar atmosphere combined in the right proportions before any reliable estimate can be made of the energy distribution there. These are not available now and are not likely to be in the near future.

In addition to continuous absorption there is the contribution of lines, which will probably be greater in the far than in the near ultra-violet. Again assuming that the far ultra-violet continuum can show them in absorption, the Lyman lines of H are likely to abstract an amount of energy of the same order as that absorbed beyond the head of the series, and the ultimate lines of such elements as He, Ne, A, N, O, Si, which lie in the far ultra-violet, will be strong, as they are known to be cosmically abundant.

Extrapolation into the far ultra-violet from the observed energy distribution in the visible and near ultra-violet by means of Planck's formula is therefore extremely risky; and our present knowledge of the mechanism of the escape of radiation through the Sun's atmosphere is not detailed enough to enable us to predict what deviation from the black-body distribution is likely to occur in the far ultra-violet.

§ 4. INDIRECT OBSERVATIONS

4.1 *Continuous emission*

Certain high-excitation solar phenomena are now widely attributed to an excess of ultra-violet radiation over the visual black-body value (*c.* 6000° K.). This tendency can probably be traced back to the Zanstra theory of planetary nebulae (1929). In this theory the stellar nuclei of these objects are supposed to emit short-wave radiation which ionizes the atoms in the nebular shell, the recombination spectra of H, He I, and He II being observable in the visible region. Photo-electrons freed by the primary process of ionization excite the low-excitation lines of [O III], [O II], and [N II], producing further visible lines. On this theory the observed strength of the Balmer lines is used to calculate the number of quanta in the far ultra-violet beyond the head of the Lyman series, and thence to deduce a "temperature" for the exciting star. But it should be noted that the temperatures deduced are really radiation temperatures at wave-lengths beyond the ionization limits of the atoms concerned. There

is no reason why a radiation appendage at these wave-lengths should not be the exciting cause; in this case the temperatures deduced will greatly exceed those observed in the visible continuum and will cease to bear their usual physical significance.

4.11 *Chromosphere and prominences*

A comparable line of reasoning may be applied to the Sun: the excitation of atomic lines in the chromosphere and prominences may be due to this mechanism of recombination after ionization by excess radiation in the far ultra-violet. Most Fraunhofer lines have an E.P. of 0 to 5 v., and there is no difficulty in explaining their excitation. Above spots, however, there is local absorption by He atoms in states of E.P. about 20 v.; and the chromosphere and prominences are at all times emitting He I and He II lines which show that the excitation level here is at least 25 v. and may reach as much as 50·8 v. Temperature radiation of 6000° K. is quite inadequate to excite these lines to the observed strength even at the low pressures existing in the chromosphere. Cillié and Menzel (1935) have shown that as regards the Balmer excitation the number of quanta beyond the Lyman limit is too low by a factor of $4{\cdot}25 \times 10^4$, and that a black-body temperature of 9800° K. is required in this region to make up the deficiency. As regards the He II line at 4686 A., observed in the chromosphere and in prominences, the efficiency of 6000° K. radiation is inadequate by a factor of the order of 10^{30}, and a temperature of 20 000° K. at 243 A. has to be postulated. It is noteworthy that in order to explain the absence of marked second-stage ionization in the metals of the chromosphere, it must be supposed that the radiation is not much enhanced at 1000–1200 A.

4.12 *Dawn flash and cometary spectra*

Observations (Slipher, 1933) of the twilight flash spectrum of the Earth's upper atmosphere give direct evidence of the excitation of molecular nitrogen to an excited level A′ (18·7 v.) of N_2^+ by solar radiation of $\lambda < 611$ A. Saha (1937) deduces that if only 10 % of the measured *F*-region electron density is produced by this ionization, the ultra-violet intensity must be about 10^6 times greater than the 6500° K. black-body value at this wave-length. There seems no alternative mechanism available, in view of the direct nature of the observation and the spectroscopic evidence as to the observed emission, and the deduced absorption is unimpeachable. Similar considerations will apply to the ionization of CO and N_2 (I.P. 14·2 v., 15·5 v.) before the emission of the resonance bands of CO^+ and N_2^+ in the tails of comets which approach close to the Sun.

4.13 *Coronal spectrum*

The question of the coronal lines is relevant here. Their origin is as yet unknown, the most plausible theory yet advanced being due to Edlén (1942), who supposes that they are forbidden transitions from metastable states in highly-ionized Fe, Ni, and Ca.* The 5303 A. line, for example, is attributable to the nebular-type transition in Fe XIV, and the 3601 A. line to a similar transition in Ni XVI, both ions being iso-electronic with neutral Al. The ground levels are ^2P with calculated separations which agree with the observed ones (18 850 $cm.^{-1}$

* See *Report* on p. 101 of this Volume.

and 27760 cm.$^{-1}$ respectively) within the error of prediction. The source of such high ionization is puzzling, but if the theory is accepted, it follows that the Sun is weak in radiation around 20–60 A.: too weak to excite these ions out of their metastable states before they can radiate the forbidden lines.

4.2 *Line emission from the chromosphere and bright eruptions*

Limited areas (20–2500 millionths) of the Sun's disk sometimes show strong line-emission in the Balmer series, the ultra-violet and infra-red Ca II lines, and He I lines of E.P. 20 v. or so (Richardson and Minkowski, 1939), the duration of the emission being 20–60 min. and its strength in Hα often exceeding that of the adjacent continuum. These chromospheric eruptions are accompanied simultaneously by short-wave radio fading on the Earth, indicating an increased ionization in the *D* layer of the ionosphere, presumably due (since only the daylight side of the Earth is affected, and the events occur simultaneously) to increased ultra-violet radiation from the Sun. The presence of the Balmer lines in emission makes it certain that H$L\alpha$ at 1215 A. must be radiated strongly from these bright flocculi, and probable that higher members of the Lyman series are also present. Similarly, the emission of λ6678 of He I ($2\,^1P$–$3\,^1D$) will certainly be succeeded by that of λ584 ($1\,^1S$–$2\,^1P$). The former line (10·16 v.) cannot ionize O_2, N_2, or O directly (12·2 v., 15·5 v., 13·56 v.), though a secondary process involving super-elastic collisions may occur (Martyn *et al.*, 1937); but the latter (21·2 v.) will ionize all constituents of the Earth's atmosphere. A rough calculation (Richardson, 1939) shows that the intensity emitted by $L\alpha$ will be expected to be 250 times that emitted by a black body at 5740° K. in the 200 A. range centred on λ1200, in which the Earth's atmosphere is relatively transparent. For a big eruption (2500 millionths of the disk) this gives a 60 % increase in radiation above the calculated value in this range of wave-lengths.

Mention should also be made of the work of Preston (1940), who suggests the possibility that X rays of wave-length 2–3 A. may be present in the solar radiation from chromospheric eruptions. Vegard (1937) goes so far as to attribute the *E* and *F* layers in the ionosphere to 12-A. X rays present in solar radiation at all times, and the abnormal ionization below the *E* layer during bright eruptions to harder X rays.

Summarizing, then, chromospheric eruptions will be expected to emit the following radiations: $L\alpha$ at 1215 A. at great strength—great enough perhaps to make the secondary ionization process referred to above assume terrestrial importance, especially at great heights, in spite of its intrinsic improbability (Preston, 1940); the other members of the Lyman series to a less extent ($L\gamma$ at 972 A. will ionize O_2 if photo-ionization to the ground state of O_2^+ can occur, which seems doubtful); the Lyman continuum below 912 A., which will ionize atomic O; the principal series of He I, $\lambda\lambda$584, 537, 522, ... to its head continuum at 506 A., all of which will ionize all the main constituents of the atmosphere and will therefore be absorbed at high levels. It is true that the eruptions are small and relatively infrequent, but their terrestrial consequences are important, and qualitatively similar bright flocculi are visible near sunspots at all times. Waldmeier (1937) classifies a range of such flocculi beginning with dark areas

and ending with the brightest eruptions. In any case the chromosphere itself is radiating all these lines continuously, though to a less extent as regards intensity, and the quasi-Lyman series of He II at λλ304, 256, ... 228 A. as well.

§ 5. CONCLUSION

Neither direct observation nor theory can at present be used to deduce the distribution of solar radiation below 2900 A. Many phenomena of widely different kinds, however, combine to suggest that the Sun radiates much more energy in the far ultra-violet than would a black body at 6000° K. Super-excitation in the solar chromosphere, in comets, and in the Earth's upper atmosphere suggests that at various wave-lengths between 1000 A. and 200 A. radiation corresponding to temperatures up to 20 000° K. is leaving the Sun. The number of wave-lengths at which the presence of such excess radiation is indicated suggests that the solar spectrum effective in producing these different phenomena is a continuum; but in addition to this, line emission in the principal series of H, He I, and He II must take place from the chromosphere at all times, and (as regards H and He I) must be multiplied many thousandfold during the transitory chromospheric eruptions. Especially marked should be the emission at 1215 A. and at 584 A., due to H and He respectively.

REFERENCES

Cillié, C. G. and Menzel, D. H., 1935. *Harv. Circ.* no. 410.
Edlén, B., 1942. *Ark. Mat. Astr. Fys.* **28** B, 1.
Greaves, W. M. H., 1939. *Observatory*, **62**, 252.
Kiepenheuer, K. O., 1937. *Z. Astrophys.* **14**, 348.
Martyn, D. F., Munro, G. H., Higgs, A. J. and Williams, S. E., 1937. *Nature, Lond.*, **140**, 603.
Meyer, E., 1941. *Helv. phys. Acta*, **14**, 625.
Meyer, E., Schein, N. and Stoll, B., 1934. *Nature, Lond.*, **134**, 535.
Preston, W. M., 1940. *Phys. Rev.* **57**, 887.
Regener, E. and H., 1934. *Phys. Z.* **35**, 788.
Richardson, R. S., 1939. *Astrophys. J.* **90**, 368.
Richardson, R. S. and Minkowski, R., 1939. *Astrophys. J.* **89**, 347.
Saha, M. N., 1937. *Proc. Roy. Soc.* A. **160**, 155.
Slipher, V. M., 1933. *Mon. Not. R. Astr. Soc.* **93**, 657.
Vegard, L., 1937. *Rep. Comm. Solar-Terr. Rel.* **5**, 136.
Waldmeier, M., 1937. *Z. Astrophys.* **14**, 91.
Zanstra, H., 1929. *Publ. D.A.O.* **4**, 209.

ABSORPTION SPECTRA AND ABSORPTION COEFFICIENTS OF ATMOSPHERIC GASES

By W. C. PRICE,
Laboratory of Physical Chemistry, Cambridge

§1. INTRODUCTION

THE object of this *Report* is to review the work on the electronic spectra and absorption coefficients of the atmospheric gases. It deals first with the absorption of the various possible constituents of the atmosphere, discussing the wave-lengths at which absorption occurs and giving a qualitative idea of the intensity of the bands by stating the path length of the gas at S.T.P. required for complete absorption of the wave-lengths under discussion. The second part deals with the problem of obtaining accurate values for the absorption coefficients throughout the range in which a substance absorbs, and in particular with the difficulties of obtaining these data in the vacuum ultra-violet region of the spectrum.

In a previous *Report* (Chapman and Price, 1936) the data on the more common constituents of the atmosphere have been reviewed. All that is necessary is to add to these data what has since been published and to correct any omissions. In particular it is necessary to deal in some detail with the spectra of the oxides of nitrogen in view of the plausible identification by Adel (1939) of N_2O and N_2O_5 bands in the infra-red Fraunhofer spectrum. Also it is desirable to consider all possible atmospheric constituents if these possess any characteristic absorption in the range of the solar spectrum observable on the earth and to see if by their spectra the amounts can be estimated or at least limited. This has been done by Sutherland and Callendar (1942–43) in the infra-red, and only the visible and near ultra-violet range will be considered here. It should also be mentioned here that two extremely useful reference books (Rollefson and Burton, 1939; Noyes and Leighton, 1941) on the spectra and photochemistry of molecules have recently appeared.

§2. ABSORPTION SPECTRA OF POSSIBLE ATMOSPHERIC CONSTITUENTS

(i) *Nitrogen*

To the data previously given on nitrogen it is necessary to add that in the 1000–800 A. region. Worley and Jenkins (1938) have confirmed the existence of a molecular Rydberg series with a photo-ionization continuum starting at 793 A. corresponding to 15·503 ev. and appearing at path length of about 0·1 mm. at S.T.P. Also in the shorter wave-length region good pictures of the Hopfield-Rydberg series converging to 661 A. (18·67 ev.) and appearing at equivalent path lengths of 0·01 mm. have been published by Takamine, Suga and Tanaka (1938). All the data indicate that the bands are relatively sharp and no dissociation continua have been obtained.

(ii) *Oxygen*

Little additional data relevant to the atmospheric problem appear to have been published on the spectrum of oxygen. Schneider (1937) has, however, shown that foreign gases have no appreciable effect on the absorption coefficient of oxygen in the Schumann region, e.g. no difference was observed by the addition of 80% nitrogen at pressures less than 15 mm. of mercury.

(iii) *Carbon dioxide*

At equivalent thicknesses of 2 mm. of the gas at S.T.P. carbon dioxide starts to absorb below 1700 A. (Rathenau, 1934). Here some very diffuse bands appear. At 1620 A. they merge into continuous absorption, and apart from some narrow regions of transparency between 1250 A. and 1170 A., this extends to all shorter wave-lengths. At smaller thicknesses (about a few tenths of a mm.) the continuous absorption below 1620 A. weakens, and between 1390 A. and 1240 A. a set of discrete bands appears. The strongest bands of CO_2 lie within the region 1150–690 A., and these give rise to intense absorption at equivalent thicknesses of only 0·01 mm. In the upper half of this region Price and Simpson (1938) found Rydberg series converging to an ionization potential of 13·73 ev. In the lower range (830–690 A.) Henning (1932) has shown that the main bands conform approximately to a doublet Rydberg series converging to an ionization potential at $18{\cdot}00 \pm 0{\cdot}02$ ev. It is clear that similarities exist between the spectra of nitrogen and carbon dioxide, and this even extends to the spectra of their ions. No quantitative work on the absorption of this gas has been undertaken.

(iv) *Water vapour*

Absorption coefficients have been measured down to about 1800 A., but this absorption at long wave-lengths is continuous in character and extremely weak. It extends to the region 1785–1550 A., where it becomes much stronger (i.e. giving complete absorption in 0·1 mm.). Below 1350 A. stronger bands appear (Rathenau, 1934; Price, 1936), some of which have been interpreted as going to an ionization potential of 12·56 ev.

(v) *Hydrogen peroxide*

This substance exhibits a weak, structureless, continuous absorption beginning between 3100–3000 A. and extending to shorter wave-lengths (Urey, Dawsey and Rice, 1929). Although the extinction coefficients have been measured, it is not possible to use them to set a limit on the hydrogen peroxide in the atmosphere because of the absence of discrete bands.

(vi) *Ozone*

This molecule is extremely important to upper atmosphere phenomena, and its absorption spectra in the visible and near ultra-violet have been very carefully measured (Colange, 1927; Chalonge and Lambrey, 1927; Läuchli, 1928, 29; Ny Tsi-Zé and Choong Shin-Piaw, 1932, 33 a, 33 b). The first absorption starts around 8000 A. and reaches a maximum at about 6000 A. The ultra-violet absorption, which is much stronger, starts around 3000 A. and reaches a maximum at 2550 A., diminishing to a low value again around 2200 A. The intermediate range has been studied by Dutheil and Dutheil (1926) and by Jakovleva

and Kondratjew (1936). An attempt has been made (Price and Simpson, 1941) to follow the spectrum of ozone into the vacuum ultra-violet, but this failed because the oxygen produced by decomposition blanketed the ozone absorption. It was, however, established that the relative transparency of ozone around 2200 A. extends down as far as 1750 A. It is quite clear from recent work that this molecule is not symmetrical, and it can, therefore, be a good emitter and absorber of radiation in both the near and far infra-red. The fact that the molecule probably has a small dipole moment associated with it, enabling it to emit radiation in the far infra-red, makes it important in connection with the temperature equilibrium in the atmosphere. It is for the same reason that the possible existence of even small amounts of the oxides of nitrogen in the atmosphere must be seriously considered. The report of the Conference on Atmospheric Ozone (*Quart. J.R. Met. Soc.*, **62**, Suppl., 1936) contains interesting information on spectroscopic and photoelectric methods of measuring the amount of ozone in the atmosphere, on the variation of the absorption coefficients with temperature, and so forth.

(vii) *Oxides of nitrogen*

N_2O. Nitrous oxide may be very important, as there may be an equivalent thickness of 1 cm. of the gas at S.T.P. present in the atmosphere according to Sutherland's interpretation of Adel's data. Only with exceedingly long path lengths and relatively high pressures can absorption in nitrous oxide be detected at wave-lengths greater than 2000 A. Because of this high transparency nitrous oxide must be relatively stable photo-chemically. This high transparency is associated with its closed-shell electronic configuration, and it resembles in this respect carbon dioxide, nitrogen and the inert gases. At high pressures and long path lengths (5 atmospheres and 33 metres) absorption of a continuous character is observed to wave-lengths as long as 3065 A. (Sponer and Bonner, 1940). Leifson (1926) finds that diffuse absorption extending from 2000–1680 A. is produced by 15 mm. of the gas at S.T.P. Duncan (1936) states that diffuse absorption extending from 2068–1768 A. is obtained with about 2 cm. of the gas (a 2-metre path length at a pressure of 7·3 mm.). This absorption is followed by a region of transparency which terminates around 1550 A., below which systems of sharp bands appear. These are much stronger than the bands in the region 2068–1768 A., which are not strong enough to give rise to appreciable absorption in 1–2 mm. of the gas. The short wave-length bands can be fitted into a plausible Rydberg series giving an ionization potential of 12·66 ev. Duncan exposed the gas to light of wave-lengths greater than 1350 A. for periods as long as 30 minutes, but was unable to detect any change in the absorption spectrum such as might arise, for instance, from the production of NO or O_3.* Direct photochemical work on nitrous oxide shows that NO, O_3 and NO_2 are among the products of photo-dissociation.

NO. This gas may be quite an important atmospheric constituent, since it is formed from N_2O by photo-dissociation and by the reaction of N_2O molecules

* In contrast to this is the case of oxygen. Ladenburg, Van Voorhis and Boyce (1932) state that with 20 minutes' exposure 75 to 80% of a volume of about 30 c.c. of oxygen at 0·27 mm. pressure had been dissociated and removed. It should, however, be stated that the different experimental conditions might have had some effect on the two experiments.

with oxygen atoms in the presence of a third body. Further. any NO_2 present in the atmosphere is likely to be converted to NO. at least during the daytime, since with high light intensities and low pressures the equilibrium $NO_2 \rightleftharpoons NO + O$ is probably very largely in favour of nitric oxide. The absorption spectrum of NO starts at 2270 A. (Leifson, 1926). The first region from 2270–1920 A. consists of a fairly sharp set of double-headed bands appearing in a thickness of about 1 mm. of the gas. Below this wave-length the bands get much stronger, after which they get progressively weaker again down to about 1400 A., where they crowd together, appearing to come to a limit in the neighbourhood of this wave-length, which corresponds to the first ionization potential of the molecule (9·1 ev.). There is no photo-ionization continuum here and little absorption to shorter wave-lengths. This latter is in accordance with theoretical expectations, since NO^+ is a closed shell like N_2, and. therefore. has no low-lying states. Flory and Johnston (1935) have analysed the bands in the neighbourhood of 2000 A. in accordance with the data of Leifson (1926) and Lambrey (1929, 30 a, 30 b). They consider that pre-dissociation occurs in the $\delta(1, 0)$ band, the primary products being nitrogen and oxygen atoms in their lowest states, $N(^4S)$ and $O(^3P)$. No typical dissociation continua or pre-dissociation bands have been found. The above results suggest a possible cycle in which nitrogen molecules combine with oxygen atoms to give nitrous oxide (probably excited ^1S oxygen atoms or metastable nitrogen molecules and the presence of a third body are necessary). This gives nitric oxide either by photochemical decomposition or by collision with an oxygen atom in the presence of a third body, and the nitric oxide reverts, at least partially, to molecular nitrogen and oxygen by photo-decomposition. A side process to this $2NO + O_2$ may give rise to some N_2O_5, for which there appears to be rather uncertain evidence in the infra-red. Thus there would be an equilibrium pressure of N_2O, NO, N_2O_5.

NO_2. Nitrogen dioxide shows strong absorption at room temperatures from 6000 A. to shorter wave-lengths (Curry and Herzberg, 1933). At 3700 A. the bands become visibly diffuse. A new set of sharp bands appears about 2550 A., but on going to shorter wave-lengths diffuseness sets in again at 2450 A. (Harris, King, Benedict and Pearse, 1940). The absorption in the region 1800–1000 A. (Price and Simpson, 1941) consists of fairly sharp bands. In the visible range the absorption coefficients have been measured by Dixon (1942). They are such as to limit the amount of nitrogen dioxide present in the atmosphere during the day to less than 0·1 mm. As has already been pointed out, photo-dissociation into nitric oxide would in any case reduce this quantity to a very low value. However, if much nitric oxide is present in the upper atmosphere then, during the night, some nitrogen dioxide might be formed which might reveal itself by Fraunhofer absorption in the first brief interval of sunrise or even in the Fraunhofer spectrum of the moon. This might indirectly put a limit on the nitric oxide and also on the nitrous oxide which is expected to produce nitric oxide by interaction with oxygen atoms. However, the formation of nitrogen dioxide from nitric oxide and oxygen atoms is termolecular and would be quite slow at low partial pressures of the products, so that a considerable amount of nitric oxide might only give rise to a small quantity of nitrogen dioxide except, perhaps, in northern latitudes in winter months.

N_2O_3. A considerable amount of work on the spectra of the various oxides of nitrogen has been done by Wulf and his collaborators at the U.S. Bureau of Chemistry and Soils. A continuous absorption lying below 2500 A., which they found in mixtures of considerable amounts of NO containing small amounts of NO_2, was attributed by them to N_2O_3 (Melvin and Wulf, 1935). When very small amounts of water were also present a group of bands starting at about 3850 A. and extending to shorter wave-lengths was obtained. These they attribute to HNO_2.

N_2O_5. The most recent work on the spectrum appears to be that of Jones and Wulf (1937), who found weak continuous absorption starting around 3800 A. and increasing rapidly towards shorter wave-lengths. Although these authors give absorption coefficients, the absence of discrete bands prevents their values from being used to set a limit on the amount of N_2O_5 possibly present in the atmosphere.

NO_3. The above authors also report an extremely strong spectrum in the region 7000–5000 A., which they attribute to a higher oxide of nitrogen, probably NO_3. Schumacher and Sprenger (1928) and Sprenger (1931) obtained a very high absorption coefficient for a maximum of this spectrum in the vicinity of 6650 A. ($\alpha_{10} = 500$ per cm. per mm. Hg). This would limit its balance in the atmosphere to considerably less than an equivalent thickness of the gas of 0·1 mm. at S.T.P.

(viii) *Hydrogen sulphide*

Continuous absorption starts around 2800 A. and extends to shorter wave-lengths. Extinction coefficients have been given by Goodeve and Stein (1931) Below 1600 A. discrete bands are found (Price, 1936) converging to an ionization potential at 10·42 ev.

(ix) *Ammonia*

Bands are obtained from 2400 A. to shorter wave-lengths (Duncan, 1935).

(x) *Saturated hydrocarbons*

All saturated hydrocarbons (CH_4, C_2H_6, etc.) are transparent to at least 1500 A., below which diffuse absorption is found.

(xi) *Formaldehyde*

Formaldehyde shows discrete bands in the range 3570–2750 A., and diffuse bands at shorter wave-lengths. The strength of the bands is such as to limit the atmospheric content to less than an equivalent thickness of 1 mm. of the gas at S.T.P. (Henri and Schou, 1928).

(xii) *Sulphur dioxide*

Sulphur dioxide absorbs strongly in the region 3160–2700 A., and an equivalent thickness of a few hundredths of a millimetre at S.T.P. is easily detected by its absorption bands around 3000 A. Hence there must be less than this amount of free sulphur dioxide in the atmosphere. Details of its absorption at shorter wave-lengths are given by Price and Simpson (1938).

(xiii) *Other possible constituents*

Details of the spectra of other possible constituents not reported here can be found by reference to the compilations of Sponer (1935), Noyes and Leighton (1941), and of Rollefson and Burton (1939).

§3. THE PROBLEM OF ABSORPTION COEFFICIENTS

It is clearly going to be important in atmospheric absorption problems to have quantitative data on the absorption curves of gases in the vacuum ultra-violet. Methods for determining the absorption curves of gases in the visible and quartz ultra-violet have been well worked out (Brode, 1939; Twyman and Allsopp, 1934). Up to the present there has been only one experimental determination of the absorption curve of a substance (Ladenburg, Van Voorhis and Boyce, 1932) in the region below 1800 A. The reason for this is that investigations in the short-wave ultra-violet necessitate working *in vacuo*, using Schumann-type plates and fluorite windows and lenses. Further, it has not been possible to use the absorption curve, obtained with so much difficulty, until recent years to test a theoretical hypothesis or even to correlate it with similar curves in other regions of the spectrum. The slight prospect of being able to interpret the results of difficult experimental work has offered little encouragement for investigations in this field. However, it has recently been possible to make important theoretical deductions from absorption curves (Mulliken, 1939), and theoretical physicists have been greatly hampered by the absence of exact quantitative data in this field. In many cases the absolute absorption coefficients provide the most reliable clue to the nature of the electronic transition. This is particularly the case in the spectra of polyatomic molecules, where rotational structure cannot be resolved and where in any case rotational and vibrational structure are difficult to interpret even in those few cases where they can be obtained. In addition to the theoretical value of quantitative absorption work it is clear that such information is vital to a satisfactory explanation of the phenomena resulting from atmospheric absorption, in photochemistry and in similar problems. There is no doubt that in the future much more attention must be given by spectroscopists to obtaining reliable quantitative information of this nature.

The above remarks are best illustrated by reference to the single publication in this field and to the valuable results that have accrued from it. Ladenburg, Van Voorhis and Boyce had made very accurate determinations of the refractive index of oxygen down to relatively short wave-lengths, and their object was to verify their dispersion formula by proving the existence of absorption bands of the required intensities in the wave-length regions calculated. In this they were successful with certain reservations, but by far the most useful results which have been forthcoming from their work are (i) the application that has been made of it to the theory of the ozone equilibrium in the atmosphere by Chapman, and (ii) its use by Mulliken in developing the theory of N→V transitions in the electronic spectra of molecules.

Ladenburg, Van Voorhis and Boyce used a small fluorite vacuum spectrograph. Their source was a hydrogen discharge tube. Interposed between the source and the spectrograph was an absorption tube the ends of which

were closed by fluorite lenses. A set of five apertures calibrated as density standards was used. Exposures were made with each of the various apertures in position and also with a known pressure of oxygen in the cell. Thus a graph could be constructed relating the microphotometer deflections at various wave-lengths with optical densities, and from this the absorption curve could be calculated.

Clearly this method is not applicable below 1350 A., where fluorite absorbs. It might be pushed to the extreme limit of 1150 A. by using lithium fluoride lenses. For wave-lengths shorter than this other methods will have to be devised. It will be necessary to use vacuum grating spectrographs, and the gas under investigation will have to be contained in the body of the spectrograph. It could be introduced by means of a gas pipette, or a constant stream could be pumped through the spectrograph. In either case helium could be used as a diluent gas, as the accurate measurement of very low pressures might otherwise be necessary. Concave mirrors could perform the focusing function of fluorite lenses. Perhaps the most serious of all the difficulties involved in the investigation is that of getting a suitable background of radiation against which the absorption can be observed and measured. The spectrum of hydrogen does not extend much below 900 A. and the gas absorbs slightly even at 1150 A. The Lyman continuum (Collins and Price, 1934) has the disadvantages that the heavy condenser discharges which have to be used liberate oxygen due to the disintegration of the glass of the capillary and also that the bore increases throughout the exposure, due to the same effect. The helium continuum (Hopfield, 1930) might be used, but it is very limited in range. Further research might enable another background to be developed, and in any case many of the other difficulties involved in this work will have to await laboratory trials. The problems involved certainly offer a wide scope for experimental skill and ingenuity.

The atmospheric absorption problem is complicated by the fact that the intensity-wave-length curve of the solar radiation between 2000 and 100 A. is likely to show enormous discontinuities, arising from absorption of elements in the Sun's chromosphere and from the line emission of hydrogen and helium (Hunter, 1942–43). If these atomic lines form a high percentage of the radiation at short wave-lengths, then the absorption coefficients of the atmospheric gases at the corresponding wave-lengths may provide useful data. A start in this direction has been made by Preston (1940), who has measured the absorption coefficients of several gases for H$L\alpha$ (1215·7 A.). He gives the following values: oxygen 0·28, nitrogen $<0{\cdot}005$, carbon dioxide 201, water vapour 390 (all values being reduced to S.T.P.). The corresponding coefficients for the higher members of the Lyman series and for all members of the helium series could probably also be determined, though with somewhat greater difficulty. Care would have to be exercised in using these results as the solar hydrogen and helium lines are wider than those from terrestrial sources, and this would lead to somewhat erroneous conclusions on the absorption of a substance with narrow absorption bands.

REFERENCES

ADEL, A., 1939. *Astrophys. J.* **90**, 627.

BRODE, W. R., 1939. *Chemical Spectroscopy* (New York : John Wiley and Sons).

CHALONGE, D. and LAMBREY, M., 1927. *C.R. Acad. Sci., Paris*, **184**, 1165.

CHAPMAN, S. and PRICE, W. C., 1936. *Rep. Prog. Phys.* **3**, 55.

COLANGE, G., 1927. *J. Phys. Radium*, **8**, 256.

COLLINS, G. and PRICE, W. C., 1934. *Rev. Sci. Instrum.* **5**, 423.

CURRY, J. and HERZBERG, G., 1933. *Nature, Lond.*, **131**, 842.

DIXON, J. K., 1940. *J. Chem. Phys.* **8**, 157.

DUNCAN, A. B. F., 1935. *Phys. Rev.* **47**, 822.

DUNCAN, A. B. F., 1936. *J. Chem. Phys.* **4**, 638.

DUTHEIL, J. and DUTHEIL, M., 1926. *J. Phys. Radium*, **7**, 415.

FLORY, P. J. and JOHNSTON, H. L., 1935. *J. Amer. Chem. Soc.* **57**, 2641.

GOODEVE, C. F. and STEIN, N. O., 1931. *Trans. Faraday Soc.* **27**, 393.

GRANATH, L. P., 1929. *Phys. Rev.* **34**, 1045.

HARRIS, L., KING, G. W., BENEDICT, W. L. and PEARSE, R. W. B., 1940. *J. Chem. Phys.* **8**, 765.

HENNING, H. J., 1932. *Ann. Phys., Lpz.*, **13**, 599.

HENRI, V. and SCHOU, S. A., 1928. *Z. Phys.* **49**, 774.

HOPFIELD, J. J., 1930. *Astrophys. J.* **72**, 133.

HUNTER, A., 1942–43. *Rep. Prog. Phys.* **9**, 5.

JAKOVLEVA, A. and KONDRATJEW, V., 1936. *Phys. Z. Sowjet.* **9**, 106.

JONES, E. J. and WULF, O. R., 1937. *J. Chem. Phys.* **5**, 873.

LADENBURG, R., VAN VOORHIS, C. C. and BOYCE, J. C., 1932. *Phys. Rev.* **40**, 1018.

LAMBREY, M., 1929. *C.R. Acad. Sci., Paris*, **189**, 574.

LAMBREY, M., 1930 a. *C.R. Acad. Sci., Paris*, **190**, 670.

LAMBREY, M., 1930 b. *Ann. Phys., Paris*, **14**, 95.

LAUCHLI, A., 1928. *Helv. phys. Acta*, **1**, 208.

LÄUCHLI, A., 1929. *Z. Phys.* **53**, 92.

LEIFSON, S. W., 1926. *Astrophys. J.* **63**, 73.

MELVIN, E. H. and WULF, O. R., 1935. *J. Chem. Phys.* **3**, 755.

MULLIKEN, R. S., 1939. *J. Chem. Phys.* **7**, 14 *et seq.*

NOYES, W. A. and LEIGHTON, P. A., 1941. *The Photochemistry of Gases* (New York : Reinhold Publishing Corporation).

NY TSI-ZÉ and CHOONG SHIN-PIAW, 1932. *C.R. Acad, Sci., Paris*, **195**, 309.

NY TSI-ZÉ and CHOONG SHIN-PIAW, 1933 a. *Chinese J. Physics*, **1**, 38.

NY TSI-ZÉ and CHOONG SHIN-PIAW, 1933 b. *C.R. Acad. Sci., Paris*, **196**, 916.

PRESTON, W. M., 1940. *Phys. Rev.* **57**, 887.

PRICE, W. C., 1936. *J. Chem. Phys.* **4**, 147.

PRICE, W. C. and SIMPSON, D. M., 1938. *Proc. Roy. Soc.* A, **169**, 501.

PRICE, W. C. and SIMPSON, D. M., 1941. *Trans. Faraday Soc.* **37**, 106.

RATHENAU, G., 1934. *Z. Phys.* **87**, 32.

ROLLEFSON, G. K. and BURTON, M., 1939. *Photochemistry and the Mechanism of Chemical Reactions* (New York : Prentice Hall).

SCHNEIDER, E. G., 1937. *J. Chem. Phys.* **5**, 106.

SPONER, H., 1935. *Molekülspektren II* (Berlin : Springer).

SPONER, H. and BONNER, L. G., 1940. *J. Chem. Phys.* **8**, 33.

SCHUMACHER, H. J. and SPRENGER, G., 1928. *Z. phys. Chem.* **136**, 77.

SPRENGER, G., 1931. *Z. Elektrochem.* **37**, 674.

SUTHERLAND, G. B. B. M. and CALLENDAR, G. S., 1942–43. *Rep. Prog. Phys.* **9**, 18.

TAKAMINE, T., SUGA, T. and TANAKA, Y., 1938. *Sci. Papers Inst. Phys. Chem. Res., Tokyo*, **34**, 854.

TWYMAN, F. and ALLSOPP, C. B., 1934. *The Practice of Absorption Spectrophotometry* (London : Adam Hilger, Ltd.).

UREY, H. C., DAWSEY, L. H. and RICE, F. O., 1929. *J. Amer. Chem. Soc.* **51**, 1371.

WORLEY, R. E. and JENKINS, F. A., 1938. *Phys. Rev.* **54**, 305.

THE INFRA-RED SPECTRA OF ATMOSPHERIC GASES OTHER THAN WATER VAPOUR

G. B. B. M. SUTHERLAND,
Pembroke College, Cambridge,
AND
G. S. CALLENDAR,
Imperial College, London

§ 1. INTRODUCTION

ALTHOUGH the existence of infra-red radiation was first discovered by Herschel in 1800, when he placed a thermometer beyond the red end of the visible spectrum of the sun, no systematic investigation of the infra-red solar spectrum was undertaken until about 1880, when Langley, using a bolometer of his own invention, began a series of brilliant researches which revealed a rich and very varied absorption spectrum. By about 1900 a very complete map was available up to nearly 5·5 μ and several of the more prominent absorptions had been satisfactorily correlated with known absorption bands of water vapour and carbon dioxide. There followed a period of relative inactivity (presumably because there were no great advances in infra-red technique) until the introduction of the echelette grating by Wood around 1920, and of new types of thermocouples by Pfund around 1926, suddenly made possible the attainment of very much higher resolving power in the infra-red. The application of the new technique to the problem of the solar infra-red spectrum has been due almost entirely to Adel.

The present-day investigator, in addition to having higher resolving power at his disposal, has the added knowledge of the infra-red spectra of practically all the possible atmospheric constituents, as these have been investigated in great detail only during the past twenty years. It should, therefore, be possible to combine these new data and lay down precise limits for the abundance of various atmospheric constituents. The present *Report* represents a first step in such an investigation, being a critical review prepared at the request of the Gassiot Committee.

§ 2. GASES KNOWN TO EXIST IN THE ATMOSPHERE

(i) *General survey*

Of the known atmospheric gases (N_2, O_2, H_2O, CO_2, O_3 and the monatomic gases) only H_2O, CO_2 and O_3 possess absorption spectra in the infra-red. Nevertheless the rich absorption spectrum of the atmosphere between the visible and 14 μ, when examined under low resolving power, can nearly all be accounted for, qualitatively at least, in terms of well-known absorption bands of these

three molecules. This is illustrated in figure 1, which indicates the positions and approximate intensities of the main bands in the spectrum, the intensities being shown as the percentage absorption of the estimated continuous background emitted by the sun. This diagram should not be treated as an accurate representation. It has been rapidly prepared from a number of sources (Langley and Abbott, 1900; Adel, Slipher and Barker, 1935; Adel, 1938), but the qualitative picture it presents is very helpful in much of the discussion to follow. Recently a much more detailed plot of the atmospheric spectrum between 2 μ and 14 μ has been made by Adel (1941 a–41 f), using a grating spectrometer. Unfortunately curves published so far by Adel (1941 c, 41 e, 41 f,) apply only to the region between 7 μ and 14 μ,* and it is not practicable to reproduce these in this *Report*.

The generally accepted interpretations of the atmospheric absorptions are marked on figure 1. It will be noticed that below 7 μ all absorptions (with the

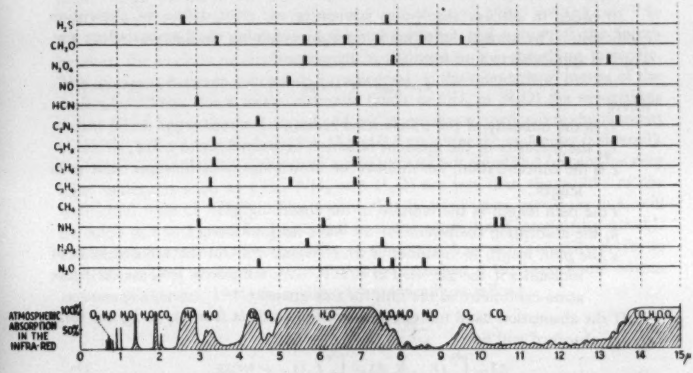


Figure 1.

exception of the electronic O_3 band at 0·76 μ) are assigned to H_2O, CO_2 and O_3; between 7 μ and 8 μ there are two weak bands at 7·6 μ and 7·8 μ originally attributed respectively to N_2O_5 (Adel and Lampland, 1938) and N_2O (Adel, 1938); at 9·6 μ is a fairly strong absorption assigned to O_3, while beyond 13 μ the absorption rises steeply to complete extinction and no appreciable energy appears to be transmitted beyond 14 μ* (Adel and Lampland, 1940; Adel, 1941 c). Several additional very weak absorptions have been recorded between 8 μ and 13 μ by Adel, of which only a few have been satisfactorily assigned. These will be discussed in the second part of this *Report*, when other possible constituents of the atmosphere come under review.

Confining our attention for the moment to the three well established constituents, H_2O, CO_2 and O_3, it is important to realise that, although all the major

* *Note added in Proof.* Adel (1942) has recently published an extension of the atmospheric spectrum to 24 μ which proves that absorption is not complete beyond 14 μ as was formerly supposed.

discrete absorption bands can be accounted for *qualitatively* in terms of the vibration-rotation spectra of these three molecules, we cannot consider the infra-red spectrum as adequately explained until the *quantitative* side has been fully investigated. Thus continuous, as well as discrete, absorption has to be taken into account, and also the possibility of other gaseous constituents, the spectra of which would be masked by the spectra of H_2O, CO_2 and O_3. The spectrum of H_2O forms the subject of a separate *Report* by Dr. Cowling, so that in this section of the *Report* we need only consider CO_2 and O_3. From the knowledge of the spectra of these molecules which has been obtained in laboratory investigations, together with the accepted estimates of the amounts of CO_2 and of O_3 in the atmosphere (based on chemical analysis and ultra-violet spectra), we shall try to predict the intensities of the atmospheric absorption bands and compare the predicted with the observed intensities.

(ii) *The relation between mean percentage absorption and gas concentration*

In order to discuss absorption intensities we must define an absorption coefficient. The normal definition is made by assuming the Lambert-Beer law, which is applicable to a uniform gas at any pressure:

$$I_\lambda = I_{0\lambda} e^{-k_\lambda cl} = I_{0\lambda} e^{-k_\lambda q}, \qquad \ldots\ldots(1)$$

where

I_λ is the intensity of the transmitted radiation between λ and $\lambda+\delta\lambda$ and

$I_{0\lambda}$ the intensity of the incident radiation between λ and $\lambda+\delta\lambda$;

c is the concentration, i.e. number of absorbing molecules per unit path length;

l the path length of the radition in the absorbing gas;

k_λ the absorption coefficient for the wave-length λ; and

q the path length in centimetres which would contain the same number of molecules if the gas were at S.T.P. We tentatively propose the name atmo-centimetre as the unit for this quantity.*

If the absorption band under consideration extends from λ_1 to λ_2, the total energy absorbed will be

$$E_1^2 = \int_{\lambda_1}^{\lambda_2} (I_{0\lambda} - I_\lambda)\delta\lambda = \int_{\lambda_1}^{\lambda_2} I_{0\lambda}(1 - e^{-k_\lambda q})\delta\lambda. \qquad \ldots\ldots(2)$$

Provided the absorption is never very high, this approximates to

$$\int_{\lambda_1}^{\lambda_2} I_{0\lambda} k_\lambda q \delta\lambda = I_{0\lambda} q \int_{\lambda_1}^{\lambda_2} k_\lambda \delta\lambda, \qquad \ldots\ldots(3)$$

if the band is sufficiently narrow for the variation in I_0 over the band to be neglected. Let $E_{0\lambda}$ be the total incident energy contained in the width of the band, then

$$E_\lambda = E_{0\lambda} q \bar{k}_\lambda,$$

where $\bar{k}_\lambda$ is the mean value of k_λ over the whole width of the band. In other words the mean percentage absorption, A, over the band is given by

$$\bar{A} = 100\bar{k}_\lambda q \qquad \ldots\ldots(4)$$

* This suggestion is due to Professor S. Chapman.

Now $\bar{k}_\lambda$ can be computed from laboratory observations, so if q is known $\bar{A}$ can be calculated or vice versa.

Before attempting to apply (4) to the atmospheric spectrum we must examine more closely the assumption that $k_\lambda q$ is small. The absorption bands with which we are dealing here are due to changes in the electric moment of the molecule when it vibrates, but when the vibrational energy changes the rotational energy will also change, giving each vibrational band a discrete rotational fine structure. This implies large variations in k_λ in traversing a band. The obvious way of overcoming this difficulty is to apply (4) only to weak individual rotation lines; unfortunately, no experimental data exist as yet for the percentage absorptions of separate rotation lines in the atmospheric spectrum. The error introduced by this "smearing out" of the rotational fine structure will obviously depend on the ratio of the line width to the mean distance between successive lines. Line widths in the infra-red have not been investigated to any extent, but reasonable limits (Herzberg, Spinks and Watson, 1936) are 0·5 cm.$^{-1}$ to 1·5 cm.$^{-1}$, depending on the pressure, temperature and nature of the forces between the various molecules present in addition to the absorbing molecule. The distances between successive lines depend on the moments of inertia of the molecule. Thus, for a linear molecule (such as CO_2 or N_2O) the separations are, to a first approximation, equal and are inversely proportional to the moment of inertia (being respectively 1·5 cm.$^{-1}$ and 0·8 cm.$^{-1}$ for the molecules instanced). For other molecules the separations are in general not equal and (except for very light molecules such as CH_4, NH_3 and H_2O) are less than the widths of the individual lines at S.T.P. This means that in many cases variations in k_λ will be small and so $k_\lambda q$ can be kept small over the whole band provided a suitable cell length is chosen. The majority of observers, however, naturally prefer to have about 50% absorption. In these circumstances it has been found advantageous (Callendar, 1941) to modify (4) empirically as follows:

$$\bar{A} = 100\frac{k_\lambda q}{1+k_\lambda q} \qquad \ldots\ldots(5)$$

$$\text{or} \quad \bar{k}_\lambda = \frac{\bar{A}}{q(100-\bar{A})}. \qquad \ldots\ldots(6)$$

Although this modification may give consistent results among a set of laboratory observations, it should not be applied to atmospheric conditions without considering the possible effects of temperature and of pressure. A rise in temperature will increase the overall absorption of a given band for the following reasons:

(i) Doppler broadening of the individual lines;
(ii) Collision broadening of the individual lines;
(iii) Additional lines and bands will appear since higher energy levels will become populated, thus permitting new transitions.

A rise in pressure will increase absorption mainly because of collision broadening. The order of magnitude of this effect will vary with the intermolecular forces which come into play. Thus it has been found to be particularly marked for HCN where the line width alters from 0·57 cm.$^{-1}$ at 130 mm. Hg

to 1·37 cm.$^{-1}$ at 650 mm. Hg (Herzberg, Spinks and Watson, 1936). An exactly similar effect will be produced by mixing a second gas with the first, so we have a third factor to take into account, viz., the effect of the presence of gas A on the absorption bands of B even when these bands do not overlap. It is not possible to calculate the alterations in k_1 arising from these three causes, and the experimental data are too scanty to allow any empirical estimates to be made. Until further experiments have been carried out we must simply use the present data, remembering, however, that these effects must impose limitations on our conclusions in a certain direction.

(iii) *Carbon dioxide and ozone*

The vibrations which give rise to the strongest atmospheric absorptions are illustrated in figure 2. The linear symmetrical CO_2 molecule has only two fundamental frequencies active in absorption (ν_2 and ν_3), whereas all three fundamentals are active in the non-linear ozone. It is interesting to note that the oxygen atoms in ozone do not lie at the corners of an equilateral triangle.

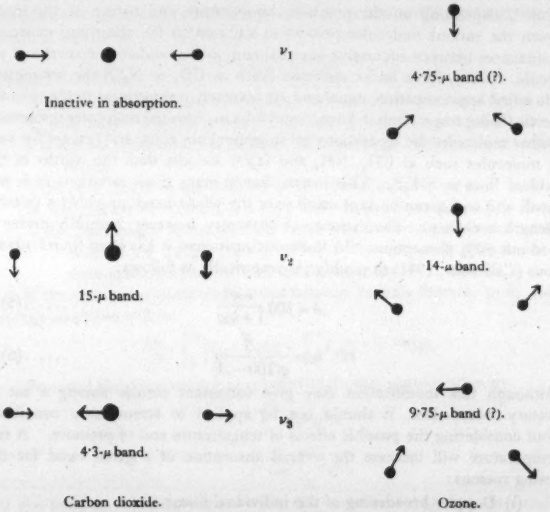


Figure 2. Fundamental modes of vibration of carbon dioxide and ozone.

If this were so, ozone would have only one fundamental vibration active in absorption.

The C—O distance in CO_2 is 1·15 A. The inter-atomic distances in ozone have not yet been determined and there is actually still some doubt here about the assignment of the fundamental frequencies and the exact value of the apical angle.

Values of k for the principal bands of CO_2 and O_3 deduced by applying (6) to the laboratory observations are given in table 1. It should be emphasized that these values are very rough, as practically no worker on infra-red spectra has troubled to measure intensities or gas pressures at all accurately: in such work the primary interest has been in establishing the positions and fine structures of absorption bands as a guide to molecular structure.

Taking the CO_2 and O_3 contents of the atmosphere as respectively 220 atmo-cm. and 0·3 atmo-cm. at S.T.P. (q in column 5), the predicted and observed percentage absorptions in the atmosphere are compared in columns 6 and 7. In the case of CO_2 the absorption due to the main bands is virtually

Table 1

Gas	Band centre in μ	Band width in μ	$\bar{k}_\lambda$	q in atmo-cms.	Predicted percentage atmospheric absorption	Observed percentage atmospheric absorption	A in per cent	E in per cent
CO_2*	4·3	4·0–4·5	2·0	220	100	100	67	0·3
	—	13–14	0·01	,,	67	60	1	5·4
	15·0	14–16	0·9	,,	100	100	47	10·0
O_3†	4·75	4·5–5·0	3·0	0·3	50	60	75	0·4
	9·65	9·4–9·8	3·0	,,	50	60	75	2·4
	14·1	12·5–15·5	0·2	,,	6	?	17	12·6

* Rubens and Ladenburg, 1905 ; Hertz, 1911 ; Schaefer and Phillips, 1926 , Martin and Barker, 1932 ; Barker and Adel, 1933.

† Gerhard, 1932 ; Hettner, Pohlmann and Schumacher, 1934.

complete, and the agreement, while satisfactory, is not very significant. However, a more detailed investigation of the short-wave side of the 15-μ CO_2 band by Callendar (1941) shows that the predicted values for 13–14 μ agree quite well with the observed values when allowance is made for the effect of lower average pressure of the gas in the atmosphere. In the case of ozone, the agreement is very satisfactory for the two bands at 4·75 μ and 9·65 μ; unfortunately, no check is possible on the band at 14·1μ since it is completely swamped by the CO_2. In the 8th column of table 1 we have put the percentage absorption A which would be observed for 1 atmo-cm., while in column 9 we have put the percentage E of the total energy of a black body at 273° K., which would fall within the wave-length range of the band. This will be useful in future applications to radiative equilibrium.

§ 3. OTHER POSSIBLE POLYATOMIC ATMOSPHERIC GASES

The following are gases the possible existence of which in the atmosphere should be considered:

Methane, CH_4
Ethylene, C_2H_4
Ethane, C_2H_6
Propane, C_3H_8
Ammonia, NH_3
Cyanogen, C_2N_2
Hydrocyanic acid, HCN
Nitric oxide, NO
Nitrous oxide, N_2O
Nitrogen dioxide, NO_2
Nitrogen tetroxide, N_2O_4
Nitrogen pentoxide, N_2O_5
Hydrogen sulphide, H_2S
Sulphur dioxide, SO_2
Formaldehyde, CH_2O.

(i) *Nitrous oxide and nitrogen pentoxide*

We shall start by discussing N_2O and N_2O_5, as Adel has brought forward evidence for their presence in the atmosphere in appreciable quantities. This evidence was originally based on the two weak bands at 7·6 μ and 7·8 μ, which were at first assigned respectively to N_2O_5 (Adel and Lampland, 1938) and N_2O (Adel, 1938). Later Adel (1939) treated the two bands as one doublet, due principally to N_2O, but having its shorter wave maximum intensified by the presence of N_2O_5. The evidence for N_2O was further increased by finding an additional weak absorption band at 8·6 μ which could also be attributed to N_2O. More recently, by using a grating spectrometer, Adel (1941 b) claims to have identified the individual rotation lines of N_2O in the atmospheric spectrum in the neighbourhood of 7·7 μ. This evidence, while very strong, is not entirely conclusive, and we have endeavoured to obtain further independent evidence in the following manner.

Laboratory observations on the 7·8-μ band of N_2O yield 0·8 as the value of k_λ for this band. From this we estimate that there are 0·8 atmo-cm. of N_2O in the atmosphere. Taking this figure in conjunction with a computed value of 0·3 for the k_λ of the N_2O band at 8·6 μ, we infer that this atmospheric band should have an average absorption intensity of 20%. Actually the intensity of the observed band appears to be less than 10%, but the uncertainties in the figures involved are too great to base any conclusions on this result taken alone.* More significant is the fact that the latest plots (Adel, 1941 f) under high resolving power show that the 8·6-μ atmospheric band has a fine structure very different from the 8·6-μ band of N_2O.

In the case of N_2O_5 the laboratory observations are very scanty (Warburg and Leithauser, 1909), and it seems to us that the evidence for its presence is rather weak. No other band of N_2O_5 has been observed in the atmospheric spectrum, and in fact the spectrum of N_2O_5 is largely unknown. The one other strong band which is known occurs at 5·8 μ, where it will be completely masked by water vapour. The main argument for N_2O_5 appears to be that in many of the early reports of the ozone spectrum in the laboratory a band was reported at 7·7 μ which was later proved (Hettner, Pohlmann and Schumacher, 1934) to be due to N_2O_5. Unfortunately, the ultra-violet spectrum lends very little support in favour either of N_2O or N_2O_5. In the case of N_2O the absorption sets in at too short a wave-length to be observed in the atmospheric spectrum, and for N_2O_5 the existing data (Urey, Dawsey and Rice, 1929; Dutta and Sen Gupta, 1933) are conflicting, and in any case are insufficient to allow an upper limit to be placed on the N_2O_5 content of the atmosphere.

Summing up, we may say that there is fairly good evidence for a few millimetres of N_2O in the atmosphere and very weak evidence for a few tenths of a millimetre of N_2O_5.

(ii) *Information available from ultra-violet spectra*

Before passing to a detailed review of the infra-red absorption due to other possible atmospheric gases, it is advisable to consider briefly whether the ultra-

* *Note added in Proof.* Dr. Adel has informed us in private correspondence that he has measured the relative intensities of the 7·8-μ and 8·6-μ bands of N_2O and that the latter band is much weaker than one would infer from the existing data.

violet or visible regions of the spectrum may not provide better evidence with regard to the upper limit of concentration of these gases in the atmosphere. It appears that for the hydrocarbons, as for HCN, NH_3, C_2N_2, N_2O_4 and H_2S, the visible and ultra-violet spectra either lie beyond the limits of atmospheric transmission or are unsuitable for other reasons. However, for NO_2 (Dixon, 1940; Daniels and Holmes, 1934), SO_2 (Clements, 1935) and CH_2O (Dieke and Kistiakowsky, 1934) safe upper limits can be set at a few hundredths of an atmo-cm. for the first two and a few tenths of an atmo-cm. for the third.* The absorption in the infra-red, due to NO_2 and SO_2, can, therefore, be neglected since amounts of that order would have but negligible effect on radiative equilibrium to that region of the spectrum. Incidentally, we might remark that with the exception of the saturated hydrocarbons none of the above gases is likely to be present in appreciable concentration near the surface of the earth except under very special conditions, but their presence in the stratosphere cannot be ruled out without further consideration.

(iii) *The remaining gases*

Following the conclusions arrived at in the foregoing section a review has been made of the spectra of the remaining gases, primarily with the object of putting an upper limit on the concentration of any of these gases in the atmosphere. The main absorption bands are marked on figure 1, and an attempt has been made to estimate the absorption coefficients (k) for as many bands as possible. These are listed in table 2, together with their reciprocals, i.e. the amount of gas in atmo-cm. required to produce 50% mean absorption in that particular region. This should provide a very safe upper limit to the amount of the gas that can possibly be present in the atmosphere. It may be added that the amount of gas in atmo-cm. multiplied by 1·4 gives the approximate parts per million by volume of the gas in the atmosphere. It will be noticed that the upper limit is surprisingly high in many cases, but it should be remembered that the k values used are mean values over a definite range of the spectrum. When k values are available relating to the strong Q branches or individual strong lines within a band, and the atmospheric spectrum has been obtained under similar conditions of high resolving power, it will be possible to lower most of these limits very considerably. In several cases it will be observed that many of the most suitable bands (i.e. those having the highest k value) coincide with intense absorptions already existing in the atmosphere due to H_2O, CO_2 and O_3. In this connection we would emphasize the importance of using high resolving power which may permit the separation of two completely overlapping bands indistinguishable under lower resolving power. From this first survey it seems unlikely that many of the new absorptions recorded by Adel between 7 μ and 14 μ can be accounted for by postulating the existence of any of the above gases, and it should be noted that Adel (1941 e) has proposed explanations for a few of these in terms of absorption lines of Na and K arising in the sun's atmosphere. On the other hand, the present examination of the spectrum is little more than cursory, and a more exhaustive comparison may yield interesting results. For instance,

* This is dealt with more fully by Dr. W. C. Price in his *Report* (p. 10). We are much indebted to Dr. Price for discussion in connection with these figures.

Table 2

Gas	Band	$\bar{k}$	$1/\bar{k}$	Remarks
NH_3 [1]	3·0	0·09	11	
	6·1	0·33	3	Overlapped by H_2O
	10·55	0·2	5	Not found in Adel's latest plot
[1] Schierkolk, 1924 ; Robertson and Fox, 1928 ; Barker, 1929 ; Dennison and Hardy, 1932.				
CH_4 [2]	3·3	0·2	5	Overlapped by H_2O (?)
	7·7	0·3	3	Overlapped by N_2O (?)
[2] Cooley, 1925 ; Nielsen and Nielsen, 1935.				
C_2H_4 [3]	3·3	0·1	10	Overlapped by H_2O
	5·3	0·1	10	,, ,, ,,
	6·9	0·1	10	,, ,, ,,
	10·5	0·2	5	Possibly present in Adel's latest plot
[3] Levin and Meyer, 1928 ; Galloway and Barker, 1942.				
C_2H_2 [4]	3·4	0·05	20	
	6·8	0·15	6·7	Overlapped by H_2O
	12·2	0·15	6·7	Not found in Adel's latest plot
[4] Levin and Meyer, 1928 ; Crawford, Avery and Linnett, 1938.				
C_2H_6 [5]	3·4	6	0·16	
	7·0	1·5	0·67	Overlapped by H_2O
[5] Wu and Barker, 1941 ; Bartholomé, 1933.				
C_2N_2 [6]	4·5	0·01	100	Overlapped by CO_2
	13·5	?	?	
[6] Burmeister, 1913 ; Stefanescu and Vencov, 1939.				
HCN [7]	3·0	0·2	5	
	7·0	0·4	2·5	Overlapped by H_2O
	14·0	3·0	0·33	,, ,, CO_2
[7] Choi and Barker, 1932.				
NO [8]	5·3	0·08	12·5	Overlapped by H_2O
[8] Snow, Rawlins and Rideal, 1929.				
N_2O_4 [9]	5·7	0·33	3	Overlapped by H_2O
	2·5	0·4	2·5	,, ,, N_2O
	13·3	0·25	4	,, ,, CO_2
[9] Sutherland, 1933.				
CH_2O [10]	3·5	0·2	5	
	5·7	0·1	10	Overlapped by H_2O
	6·7	0·04	25	,, ,, ,,
	7·5	0·02	50	Edge of H_2O
	8·6	0·02	50	Overlapped by N_2O
	9·6	0·02	50	,, ,, O_3
[10] Patty, 1932 ; Nielsen, 1934 ; Ebers and Nielsen, 1937.				
H_2S [11]	2·64	?	?	Overlapped by H_2O
	7·75	0·03	33	,, ,, N_2O
[11] Bailey, Thompson and Hale, 1936.				

there are quite a number of resemblances between the atmospheric spectrum and those of particular gases such as methane, near 7·7 μ; ethylene, near 10·5 μ; formaldehyde, between 8 μ and 12 μ; and so on. These may be merely coincidences, but a better plot of the atmospheric spectrum near 3·3 μ would be of immense value in settling many of these questions. The atmospheric absorption in this region is always attributed entirely to water vapour (overtone of the strong fundamental at 6·3 μ), but it is quite possible that higher resolving power might reveal spectra of other molecules.

§ 4. CONCLUSIONS

It should now be clear that a vast amount of work, largely experimental, is required before the problems of radiative equilibrium can be solved. The following seem to us the more urgent requirements:

(*a*) Accurate measurements on the absolute absorption coefficients in the principal bands of CO_2, O_3 and N_2O at a series of different pressures and temperatures and in the presence of varying amounts of water vapour.

(*b*) Accurate measurements of the absolute absorption coefficients for at least one band of each of the following: NH_3, CH_4, C_2H_4, C_2H_2, C_3H_8, C_2N_2, HCN, N_2O_4, N_2O_5. The most suitable bands can be decided by referring to figure 1 and (*d*) below.

(*c*) Accurate plot of the atmospheric spectrum between 2 μ and 7 μ, using a slit width of 1 cm.$^{-1}$ or less on a grating spectrometer.*

(*d*) Estimates of the relative importance of various absorption bands of the known and the fairly probable atmospheric gases, in the absorption and emission of radiation. This investigation is being started now, even though it cannot be completed until the accurate values of (*a*) and (*b*) are available. Such an investigation will, in fact, assist in indicating which gases and which bands should come first in the experimental programme.

REFERENCES

Adel, A., 1938. *Astrophys. J.* **88**, 186.
Adel, A., 1939. *Astrophys. J.* **90**, 627.
Adel, A., 1941 a. *Phys. Rev.* **59**, 915.
Adel, A., 1941 b. *Astrophys. J.* **93**, 509.
Adel, A., 1941 c. *Astrophys. J.* **94**, 375.
Adel, A., 1941 d. *Astrophys. J.* **94**, 379.
Adel, A., 1941 e. *Astrophys. J.* **94**, 449.
Adel, A., 1941 f. *Astrophys. J.* **94**, 451. (We have indicated a few of the more prominent new bands revealed by this plot in figure 1.)
Adel, A., 1942. *Astrophys. J.* **96**, 239.
Adel, A. and Lampland, C. O., 1938. *Astrophys. J.* **87**, 198.
Adel, A. and Lampland, C. O., 1940. *Astrophys. J.* **91**, 481.
Adel, A., Slipher, V. M. and Barker, E. F., 1935. *Phys. Rev.* **47**, 582.
Bailey, C. R., Thompson, J. W. and Hale, J. B., 1936. *J. Chem. Phys.* **4**, 625.
Barker, E. F., 1929. *Phys. Rev.* **33**, 685.
Barker, E. F. and Adel, A., 1933. *Phys. Rev.* **44**, 185.
Bartholomé, E., 1933. *Z. phys. Chem.* B, **33**, 152.

* *Note added in Proof.* Dr. Adel has informed us in private correspondence that he has carried out such an investigation, but that pressure of more urgent work has so far prevented publication.

BURMEISTER, E., 1913. *Verh. dtsch. phys. Ges.* **15**, 589.
CALLENDAR, G. S., 1941. *Quart. J.R. Met. Soc.* **67**, 263.
CHOI, K. N. and BARKER, E. F., 1932. *Phys. Rev.* **42**, 777.
CLEMENTS, J. H., 1935. *Phys. Rev.* **47**, 224.
COOLEY, J. P., 1925. *Astrophys. J.* **62**, 73.
CRAWFORD, B. M., AVERY, J. H. and LINNETT, J. W., 1938. *J. Chem. Phys.* **6**, 682.
DANIELS, F. and HOLMES, H. H., 1934. *J Amer. Chem. Soc.* **56**, 630.
DENNISON, D. M. and HARDY, J. D., 1932. *Phys. Rev.* **39**, 938.
DIEKE, G. and KISTIAKOWSKY, G. B., 1934. *Phys. Rev.* **45**, 4.
DIXON, J. K., 1940. *J. Chem. Phys.* **8**, 157.
DUTTA, A. K. and SEN GUPTA, P. K., 1933. *Proc. Roy. Soc.* A, **139**, 397.
EBERS, E. S. and NIELSEN, H. H., 1937. *J. Chem. Phys.* **5**, 822.
GALLAWAY, W. S. and BARKER, E. F., 1942. *J. Chem. Phys.* **10**, 88.
GERHARD, S. L., 1932. *Phys. Rev.* **42**, 622.
HERTZ, G., 1911. *Verh. dtsch. phys. Ges.* **13**, 617.
HERZBERG, G., SPINKS, J. W. and WATSON, W., 1936. *Phys. Rev.* **50**, 1186.
HETTNER, G., POHLMANN, R. and SCHUMACHER, H. J., 1934. *Z. Phys.* **91**, 372.
LANGLEY, S. P. and ABBOTT, C. G., 1900. *Ann. Geophys. Obs. Smithsonian Inst.*
LEVIN, A. and MEYER, C. F., 1928. *J. Opt. Soc. Amer.* **16**, 137.
MARTIN, P. E. and BARKER, E. F., 1932. *Phys. Rev.* **41**, 291.
NIELSEN, A. H. and NIELSEN, H. H., 1935. *Phys. Rev.* **48**, 864.
NIELSEN, H. H., 1934. *Phys. Rev.* **46**, 117.
PATTY, J. R., 1932. *Phys. Rev.* **39**, 957.
ROBERTSON, R. and FOX, J. J., 1928. *Proc. Roy. Soc.* A, **120**, 161.
RUBENS, H. and LADENBURG, E., 1905. *Verh. dtsch. phys. Ges.* **7**, 170.
SCHAEFER, C. and PHILLIPS, B., 1926. *Z. Phys.* **36**, 641.
SCHIERKOLK, K., 1924. *Z. Phys.* **29**, 277.
SNOW, C. P., RAWLINS F. I. G. and RIDEAL, E. K., 1929. *Proc. Roy. Soc.* A, **124**, 453.
SPRAGUE, A. D. and NIELSEN, H. H., 1937. *J. Chem. Phys.* **5**, 85.
STEFANESCU, D. and VENCOV, S., 1939. *Soc. Romana Phys.* **40**, 25.
SUTHERLAND, G. B. B. M., 1933. *Proc. Roy. Soc.* A, **141**, 342.
UREY, H. C., DAWSEY, L. H. and RICE, F. O., 1929. *J. Amer. Chem. Soc.* **51**, 3190.
WARBURG, E. and LEITHAUSER, G., 1907. *Ann. Phys., Lpz.*, **23**, 209.
WARBURG, E. and LEITHAUSER, G., 1909. *Ann. Phys., Lpz.*, **28**, 313.
WU, V. L. and BARKER, E. F., 1941. *J. Chem. Phys.* **9**, 487.

THE ABSORPTION OF WATER VAPOUR IN THE FAR INFRA-RED

T. G. COWLING,
The University, Manchester

§1. INTRODUCTION

In the following *Report* we shall consider the absorption of water vapour as it affects physical meteorology. The range of wave-lengths studied is chiefly that beyond 5 μ. The determination of the positions of absorption lines is not studied in detail. Attention is focused on the total absorption in fairly large wave-length ranges, and on the variation of this with temperature, pressure and the amount of absorbing vapour.

The infra-red absorption of gases and vapours was studied by many physicists prior to 1920. Their pioneer work established many of the general properties of the absorption. Later, the attention of physicists was diverted to the determination of line positions, as giving more information about molecular structure. In consequence, several of the questions left unanswered in the earlier work still remain to be answered. Latterly, interest in absorption problems has revived. The total absorption of water vapour, integrated over all wave-lengths, has been studied, and attempts have been made to determine the absorption from line intensities and line positions. Nevertheless, our information is still sadly incomplete, and suggestions for further work are therefore included.

§2. THE WATER MOLECULE

The molecules of water vapour consist of three atoms arranged at the corners of an isosceles triangle, as shown in figure 1.

The OH distance in the unexcited molecule is $0{\cdot}958 \times 10^{-8}$ cm., and the angle between the two OH lines is 104° 31′. The principal axes of inertia at the mass centre G are GA, GB in the plane of the molecule, and GC normal to this plane. The moments of inertia about these lines are $I_A = 1{\cdot}007 \times 10^{-40}$, $I_B = 1{\cdot}929 \times 10^{-40}$, and $I_C = 3{\cdot}013 \times 10^{-40}$ g.cm^2, i.e., are roughly in the ratio 1 : 2 : 3 (Darling and Dennison, 1940). They have rather different values in the excited states, when vibrational displacements and centrifugal stretching occur (Mecke, 1933). The molecule has a permanent electric dipole moment, which in the unexcited state is directed along GB and equals $1{\cdot}84 \times 10^{-18}$ e.s.u.

In liquid water, molecules tend to lose their individuality. Aggregates tend to form, composed of several molecules loosely linked together; also ions like H^+, OH^-, OH_3^+ (Bernal and Fowler, 1933). The aggregates are less likely to form when the water is diluted with certain other liquids. Similar aggregates may possibly occur in saturated water vapour or steam, especially when not mixed with other gases. In the atmosphere, aggregation is likely to be unimportant, save in cloudy regions; but its possibility in some laboratory experiments with pure vapour cannot be ignored.

§3. INTERNAL MOTIONS

Any molecule can rotate as a whole, and its atoms can also vibrate relative to each other. According to classical dynamics, the possible rotations are such that either GA or GC (figure 1) describes a (non-circular) cone round the axis of angular momentum, and in addition the molecule turns about GA or GC. When GA describes a cone of small angle, the motion, being nearly a rotation

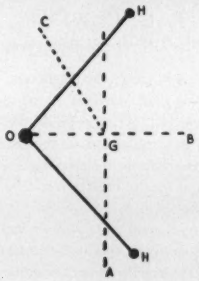
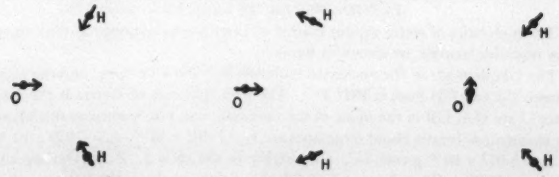


Figure 1.

about GA, is not greatly affected by the inequality of the moments of inertia I_B and I_C; similarly for I_B and I_A, when GC describes a cone of small angle. The possible internal vibrations are roughly as shown in figure 2. The oscillations ν_1 and ν_2 are symmetric, ν_3 is unsymmetric. The frequencies ν_1, ν_2, ν_3 are only to a first approximation independent of the amplitude (Dennison, 1931).



ν_1-frequency. ν_2-frequency. ν_3-frequency.

Figure 2.

On the quantum theory, only discrete rotational and vibrational motions are permitted. The possible vibrational energies are

$$h\{(n_1+\tfrac{1}{2})\nu_1+(n_2+\tfrac{1}{2})\nu_2+(n_3+\tfrac{1}{2})\nu_3\},$$

where n_1, n_2, n_3 are positive integers or zero; ν_1, ν_2, ν_3 vary slightly with n_1, n_2, n_3. The rotational states are fixed by two quantum numbers J, τ. Of these, the first, J, fixes the total angular momentum; the second, τ, which can range from $-J$ to $+J$, has no simple meaning. For given J, the total energy increases with τ. The states near $\tau=-J$ correspond to classical rotations in which GC

describes a cone of small angle, and form a series of close pairs—the energies for a pair would be equal if I_A were equal to I_B. The states near $\tau = +J$ similarly form close pairs; they are rotations in which GA describes a cone of small angle. The intermediate states are less regularly spaced.*

§ 4. ABSORPTION AND EMISSION

A molecule can absorb or emit radiation if it passes by a permitted transition from one quantum state to another. In problems of atmospheric absorption, most molecules are in the ground state for vibrations ($n_1 = n_2 = n_3 = 0$), and we need only consider transitions from this state. If the rotational motion alone alters, we get lines of the rotation band, beginning (for absorption in the atmosphere) about 15 μ, and extending to the remote infra-red. About 200 lines of this are known (Randall, Dennison, Ginsburg and Weber, 1937). If both the rotational and the vibrational states alter, we get lines of one of a series of vibration bands, clustering round the frequencies $N_1\nu_1 + N_2\nu_2 + N_3\nu_3$ (N_1, N_2, N_3 are positive integers or zero). Seventeen of these bands are known; they range from the visible region to one with its centre near 6·3 μ. Only this last band ($N_1 = 0 = N_3$, $N_2 = 1$) needs consideration here. This is a double band, i.e., it has few and weak lines near the band centre, and a pronounced minimum of absorption there. Many lines of this band are also known (Nielsen, 1941).

The theoretical determination of the line positions is difficult, because of the centrifugal stretching of the molecule. This increases the principal moments of inertia of a rapidly rotating molecule, as a consequence of which the higher rotational energy levels lie closer together. The exact calculation of the line intensities is yet more difficult. However, if I_B were equal to I_A or I_C, the intensities would be given by known formulae; and since rotational states near $\tau = -J$ are little affected by $I_A - I_B$, and states near $\tau = +J$ are little affected by $I_B - I_C$, the formulae for the cases $I_B = I_A$ or $I_B = I_C$ may be expected to give good approximations to the intensity for certain lines. Other intensities may then be found by interpolation. The intensities thus calculated for the rotation band show general agreement with those observed (Randall, Dennison, Ginsburg and Weber, 1937). They are, however, subject to systematic errors due to centrifugal stretching, etc.; and certain lines forbidden to a symmetric rotator ($I_B = I_A$ or $I_B = I_C$) are actually observed.

§ 5. LINE BREADTHS

An absorption line is not infinitely sharp; absorption occurs throughout a range of frequencies round a mean position. The mass-absorption coefficient k_ν in this range is usually supposed to be given by the Lorentz formula (Margenau and Watson, 1936),

$$k_\nu = k\gamma\pi^{-1}/\{(\nu - \nu_0)^2 + \gamma^2\}.$$

In this, 2γ represents the line width; if $\nu = \nu_0 \pm \gamma$, the absorption coefficient is half that at the line centre $\nu = \nu_0$. In the classical derivation of this formula, the width is supposed to result from molecular collisions, $1/2\pi\gamma$ being the time between successive collisions. According to quantum ideas, the width is due

* It is sometimes incorrectly stated that in these intermediate states the motion is nearly one of rotation about GB. In fact, no motions exist which are nearly rotations about GB.

to the perturbation of one molecule by another at a collision, whereby it becomes able to absorb a frequency slightly different from ν_0. The difference between the energy $h\nu$ of the radiation and the energy $h\nu_0$ required to alter the molecule's quantum state is supplied from the translatory energy of the colliding molecules; thus the collision is inelastic or hyperelastic.

The Lorentz formula cannot be exact. The classical derivation assumes no interaction between molecules save at collisions, which is not exactly realized. Again, large changes in translatory energy occur only if molecules penetrate deeply into each other's molecular fields at encounter, and in this case the details of the encounter must be considered. These details are ignored in deriving the formula, which accordingly cannot be exact when $\nu - \nu_0$ is large. Moreover, the absorption during a given collision depends on the initial velocities and internal motions of both colliding molecules; hence the use of a fixed γ for a given line is only approximately valid.

Collisions with like molecules in general give a larger γ than collisions with unlike molecules, and so γ is larger in pure water vapour than in damp air at the same total pressure. Since γ is inversely proportional to the collision interval, it is, for a gas of given composition and temperature, proportional to the pressure. If γ is due to collisions with like molecules, it is (at constant pressure) roughly inversely proportional to the temperature T; if to collisions with unlike molecules, it varies roughly as $T^{-7/10}$.

The measured value of γ for atmospheric water-vapour lines, in bands near $1\,\mu$, is about $0{\cdot}25\,\text{cm}^{-1}$ (Cornell, 1937). Its value for bands at greater wave-lengths should be of the same order of magnitude; but a direct estimate is difficult since, even in the most recent experiments on these bands, the slit widths correspond to a frequency range of order $0{\cdot}3\,\text{cm}^{-1}$. The value $\gamma = 0{\cdot}25\,\text{cm}^{-1}$ corresponds to a line width of $0{\cdot}02\,\mu$ at $20\,\mu$, and of $0{\cdot}002\,\mu$ at $6\,\mu$.

§6. ABSORPTION BY A LINE SPECTRUM

Consider the absorption when radiation from a source passes through a layer of damp air containing a thickness w cm. of precipitable water, i.e., vapour which, if precipitated as liquid water, would form a layer w cm. thick. If k_ν is the mass-absorption coefficient, the radiation of frequency ν is reduced in the ratio $e^{-k_\nu w}$. Near the centre of a strong absorption line, k_ν is so large that absorption is nearly complete for air thicknesses of the order of a few cm. ($w \sim 10^{-4}\,\text{cm.}$: Wright and Randall, 1933). As the air thickness increases above such values, less strong lines and the wings of the stronger lines produce the increase of absorption. When w is large, the central parts of all of the lines are wholly blacked out, and increased absorption can occur only in regions far from the line centres ($(\nu - \nu_0)/\gamma$ large). Since here k_ν is small, the absorption increases very slowly with increasing w. Thus, if we consider a range including several lines, we cannot find a logarithmic absorption coefficient $\bar{k}$ such that the radiation transmitted in this range is proportional to $e^{-\bar{k}w}$. If we equate the transmitted fraction of the radiation to $e^{-\tau}$, then $d\tau/dw$ falls off steadily as w increases. Use of the expression $e^{-\bar{k}w}$ must lead to wholly inaccurate results unless we are prepared to consider spectral ranges of order $0{\cdot}001\,\mu$, say. This would clearly be impracticable in meteorological work, even if experiments gave all the desired data.

To see how the total absorption actually varies with w, consider the absorption of radiation of intensity I_ν in a single isolated line (Ladenburg and Reiche, 1913). Using the Lorentz formula for k_ν, this is

$$\int I_\nu \left\{1 - e^{-\frac{k\gamma w}{\pi[(\nu-\nu_0)^2+\gamma^2]}}\right\} d\nu.$$

If w is so small that there is little absorption, even at the line centre, this is approximately

$$I_{\nu_0} \int \frac{k\gamma w d\nu}{\pi[(\nu-\nu_0)^2+\gamma^2]} = I_{\nu_0} k w,$$

i.e., it increases linearly with w. As w increases, total extinction is reached near the line centre. The absorption then becomes approximately

$$I_{\nu_0} \int \left\{1 - e^{-\frac{k\gamma w}{\pi(\nu-\nu_0)^2}}\right\} d\nu = 2 I_{\nu_0} (k\gamma w)^{\frac{1}{2}},$$

and so varies like $(\gamma w)^{\frac{1}{2}}$.

The average spacing between water-vapour lines is fairly large compared with the line width. Hence in a range $\Delta\nu$ which, though small, is large enough to contain several lines, the absorbed fraction of the radiation is

$$\frac{w}{\Delta\nu} \Sigma k$$

if w is very small, and

$$\frac{2}{\Delta\nu} (\gamma w)^{\frac{1}{2}} \Sigma k^{\frac{1}{2}}$$

for larger values of w.[*] Proportionality with $w^{\frac{1}{2}}$ is actually found when the absorption varies between wide limits—roughly 10 % to 50 % of the incident radiation (Strong, 1941). This gives an experimental check on the validity of the Lorentz formula, on which the above theory is based. Because of the variation of opacity with frequency, the range of w in which the $w^{\frac{1}{2}}$ law applies also varies with frequency, and so the $w^{\frac{1}{2}}$ law does not apply with the same exactness to the total radiation, integrated over all wave-lengths. Nevertheless, the emission of radiation by air, and atmospheric absorption of black-body radiation of ordinary temperatures, roughly follow the $w^{\frac{1}{2}}$ law when the fraction absorbed is between about 10 % and 25 % of the corresponding black-body radiation.

The $w^{\frac{1}{2}}$ law remains valid as w increases until, because of the increase of absorption in regions far from the line centres, it becomes impossible to regard the lines as isolated from each other. Thereafter the absorption increases less rapidly than $w^{\frac{1}{2}}$ (Elsasser, 1938 c; Schnaidt, 1939). The fractional absorption in the range $\Delta\nu$ becomes

$$(\Delta\nu)^{-1} \int \{1 - e^{-\gamma w \Sigma k \pi^{-1} (\nu-\nu_0)^{-2}}\} d\nu,$$

and so is given by a complicated function of γw. For a set of equal, and equally spaced, lines the fractional absorption between two line centres is

$$\frac{1}{\delta} \int_0^\delta \{1 - e^{-\gamma w k \pi^{-1} \Sigma (x-r\delta)^{-2}}\} dx$$

$$= \frac{1}{\delta} \int_0^\delta \{1 - e^{-\gamma w k \pi \delta^{-2} \operatorname{cosec}^2 (x\pi/\delta)}\} dx$$

$$= \operatorname{erf}\left(\frac{1}{\delta} (\gamma k w \pi)^{\frac{1}{2}}\right),$$

where δ is the frequency-difference between successive lines, and $\mathrm{erf}(x)$ is the error function defined by

$$\mathrm{erf}(x) = \frac{2}{\pi^{\frac{1}{2}}} \int_0^x e^{-t^2} dt.$$

This indicates that the last remnants of the light transmitted through a large mass of water vapour are extinguished roughly exponentially as w increases.

§ 7. VARIATION WITH PRESSURE

The line width 2γ does not vary greatly with the partial pressure of water vapour in atmospheric air. Thus the total absorption, at a given temperature and pressure, depends only on the total amount w of water vapour traversed, and not on the vapour pressure (Schnaidt, 1937).

The absorption does, however, vary markedly with the total pressure p, since $\gamma \propto p$. By the results of § 6 the absorption should be independent of p when w is very small, and proportional to $p^{\frac{1}{2}}$ for larger values of w; and should vary more slowly than $p^{\frac{1}{2}}$ when the absorption becomes almost complete. The proportionality with $p^{\frac{1}{2}}$ is shown in v. Bahr's observations (1909) at $2{\cdot}7\,\mu$.* On the other hand, Kühne's (1933) results between $43\,\mu$ and $60\,\mu$ indicate a slower increase with pressure than would be expected from theory.

For any special value of w, absorption in some wave-lengths will be small and not vary with p; that in others will be large, and vary more slowly than $p^{\frac{1}{2}}$. Hence the total absorption, integrated over all wave-lengths, can never increase as rapidly as $p^{\frac{1}{2}}$. Falckenberg's work (1938) has been interpreted as showing that the absorption, when about 15 % complete, varies like $p^{\frac{1}{2}}$; but this probably under-estimates the variation.

The variation of absorption with pressure is of special importance in considering the heat balance of the upper atmosphere.

§ 8. VARIATION WITH TEMPERATURE

As noted earlier (§ 5), γ varies roughly like $T^{-7/10}$ when the pressure is fixed. This relation is not likely to be followed very closely by any one line; but in any case it suggests that the variation in *line width* due to changes of temperature in the atmosphere is much less important than that due to changes of pressure.

Temperature changes do, however, affect the total absorption in certain ranges to a large extent. The major part of the absorption in the rotational band at wave-lengths less than $25\,\mu$ is produced by molecules with large rotational energies; the same is true of absorption on the flanks of the $6{\cdot}3\,\mu$ band. Thus the effect of an increase in temperature is to widen both of these bands, since it produces a big increase in the number of molecules with large rotational energies.

Boltzmann's law states that the ratio of the number of molecules in a rotational state with wave-number n (energy hcn) to the number in the ground state is proportional to $e^{-an/T}$, where a is a constant roughly equal to 1·5 cm./deg. Many lines near $18\,\mu$ arise from absorptions in which the initial wave-number is about 2000 cm.$^{-1}$. For these, $e^{-an/T}$ is twelve times as large when $T = 400°$ as when

* v. Bahr's results do not show proportionality with $p^{\frac{1}{2}}$ for low pressures. The reason probably is that the increase in pressure was secured by adding quantities of a foreign gas to a fixed quantity of saturated water vapour. The measurements at the lowest pressure refer to pure water vapour, in which γ is probably much larger (see § 9).

$T = 300°$ K. The actual strength of lines near 18 μ is not increased in this ratio, because the number of molecules in the ground state decreases as T increases; but a six-fold increase is probable. The increase at 12 μ is probably more like 100-fold.

Temperature changes in the absorption chiefly affect wave-lengths where water vapour is not very opaque; the changes are small above 0° C. at wave-lengths beyond 35 μ, and between 5·4 μ and 7·5 μ. Hence the total absorption or emission of a small quantity of water vapour ($w < 0{\cdot}005$ cm., say) will not be greatly affected by the temperature changes. The effects are much more pronounced when a few millimetres of precipitable water are present.

The temperature variation of absorption is important in problems of the upper atmosphere. It is also important in interpreting experiments on steam absorption.

§ 9. STEAM ABSORPTION

It has long been known that absorption bands in steam are much wider than in damp air. Hettner's (1918) diagram for steam shows the 6·3 μ band much wider than v. Bahr (1913) found in air; it also shows the rotation band as continuing down to 11 μ, whereas Fowle (1917) found little absorption below 13 μ in an air column containing a greater thickness of water. Weber and Randall (1932) also found a much greater absorption for steam than for damp air at 16·5 μ.

But steam absorption is also greater than that of damp air in the interior of the bands, where temperature effects should be slight. Fowle (1917) found that 0·008 cm. of water in air produces very little more absorption near the centre of the 6·3 μ band than 0·004 cm. of water as steam. Schmidt and Eckert (1936) found that the total emission (integrated over all wave-lengths) of steam diluted with a large amount of nitrogen is little more than half that of an equal amount of pure steam at the same temperature and pressure. This can be explained by supposing that the line width 2γ is much greater in pure steam than in diluted water vapour. Cornell (1937) found line widths near 1 μ which are not much greater in steam than in air; but he found a decided variation from band to band in air—in the 0·94 μ band $2\gamma = 0{\cdot}57$ cm^{-1}, and in the 1·135 μ band $2\gamma = 0{\cdot}4$ cm^{-1}. Thus a larger difference between steam and air may well be found in bands at greater wave-lengths.

For these reasons it is not safe to use steam absorptions in atmospheric problems. This is unfortunate, as a large part of the experimental data refers to steam.

§ 10. OBSERVATIONS OF TOTAL ABSORPTION

Absorption measurements can be divided into two classes, those for the separate wave-length ranges, and those of total absorption, integrated over all wave-lengths. The observations of total absorption are useful to check the accuracy of data derived for the separate wave-lengths. We consider these observations first.

In each set of observations, radiation from a black-body source is passed through a layer of damp air. The radiation measured comes partly from the source, partly from the air itself. If the source is very cold, effectively the emission of the air alone is measured; this known, the absorption of radiation from a hot source can be determined. The absorption depends on the source

temperature T_s and the air temperature T_a; it is equal to the air emission when $T_s = T_a$. Considered as a fraction $f(T_s, T_a)$ of the black-body emission at temperature T_s, it increases with T_a, but decreases with increasing T_s, because of the increased importance of shorter wave-lengths, where water vapour is less opaque. The fractional emission $f(T_a, T_a)$ increases slowly with T_a, at least for small water content.

The principal experiments on total absorption are those made by Falckenberg (1936–39), Elsasser (1941) and Brooks (1941). In Falckenberg's first paper, the source temperatures were 100° c., 40° c. and 0° c., and the difference between $f(T_s, T_a)$ and $f(T_a, T_a)$ was ignored; as Schnaidt (1937) showed, the error thus introduced is by no means small. In Falckenberg's 1939 work, and in that of Elsasser and Brooks, air emission is determined by observations with a cold source (temperature about −190° c.). The thickness w of precipitable water ranged, in Falckenberg's and Brooks's work, up to about 0·003 cm., and in Elsasser's, up to 0·3 cm. The results have to be corrected for CO_2 emission. Brooks obtained results for larger amounts of water by observations of the emission of the whole atmosphere at different zenith distances; these include ozone emission as well as that of CO_2.

The values of the fractional emission $f(T_a, T_a)$ given by different workers do not agree very well. Table 1 gives *very roughly* the values at room temperature, corrected, where necessary, by subtracting the estimated CO_2 emission, for different values of w.

Table 1

w (cm.) =	10^{-4}	10^{-3}	$10^{-2.5}$	10^{-2}	10^{-1}	$10^{-0.5}$
$f(T_a, T_a)$ =	2%	8%	16%	25%	42%	50%

These figures agree with the supposition that absorption is chiefly confined to wave-lengths beyond 25 μ, or between 5·5 μ and 7·5 μ, if $w < 10^{-3}$ cm., but that absorption in all wave-lengths beyond 17 μ is becoming large when $w \sim 10^{-1}$ cm.

In addition to the above, there are numerous other observations of sky emission. These require correction for both CO_2 and ozone emission, and we do not consider them in detail. They agree roughly with the above table for $w > 1$ mm. It is interesting also to note that experiments on water-vapour emission and absorption have been made for engineering purposes by Schmidt and Eckert (1936) and Hottel and Mangelsdorf (1935); the latter also considered the emission of CO_2. Their results apply chiefly to high temperatures, and show clearly the variation of emission with air temperature.

§ 11. OBSERVATIONS FOR THE SEPARATE WAVE-LENGTHS

We next consider the absorption at different wave-lengths. In view of the conclusions of § 9, we shall consider only experiments with damp air. The optical resolution in most experiments is small, so that no useful results could be obtained by considering extremely small wave-length intervals.

The most important set of laboratory experiments is that of Fowle (1917). Fowle measured the absorption in columns of air up to 200 m. long, the value of w ranging from 0·0035 to 0·25 cm. The mean absorption in intervals from 0·45 μ to 1 μ long was found in the range 1·3–15 μ. With 0·08 cm. of water, absorption is more than 90 % complete between 5·6 μ and 7 μ, 40 % at 5 μ, and

15 % at 8 μ. There is little absorption between 9 μ and 13·5 μ, apart from CO_2 absorption near 13·5 μ.

Because of the nearly complete transparency of short atmospheric columns beyond 9 μ, Fowle also studied the transparency of the atmosphere to sunlight in this range. His work has since been extended by Adel and his co-workers. The absorption in this range includes, besides water-vapour absorption, that of a fairly strong ozone band at 9·6 μ and a very strong CO_2 band centred at 15 μ. The latter cuts off the sun's radiation completely beyond about 14 μ, the precise position of the cut-off varying with the humidity and the sun's zenith distance.

Observations by Aldrich (quoted by Fowle) and by Adel, Slipher and Barker (1935) fail to reveal any appreciable transmission on the long-wave side of the CO_2 band (beyond 17 μ); this is interpreted as due to absorption by the rotational band of water vapour. Since the sun emits very little light beyond 17 μ, a small transmission might occur without detection; but laboratory experiments by Strong (1941) with 7 m. of damp air suggest that even at 18–19 μ, where there is a maximum of transmission, absorption is comparable with, or greater than, that in the range 13–14 μ, where it is large.*

Adel's work extends down to 5·5 μ, and his results for the 6·3 μ band serve to supplement those of Fowle. He finds that the band produces complete absorption between 5·5 μ and 7·4 μ for 1 cm. of absorbing water, while, for 1 mm. water, absorption is complete only between 5·7 μ and 7 μ, and there is a "window" at 6·3 μ in the middle of this range, transmitting a little light (Adel and Lampland, 1940a). Fowle's results can be further supplemented by those of v. Bahr (1913) on the absorption of 220 cm. of moist air between 4·8 μ and 7·5 μ. The optical resolution in v. Bahr's work is much better than in Fowle's; but the water thickness was not given by v. Bahr, and has to be found by a comparison with Fowle's results.

Between the 6·3 μ band and the CO_2 cut-off, Adel has analysed the absorption both with a prism and with a grating. The prismatic spectrum shows numerous absorption lines, many of which can be identified as due to water-vapour absorption † (Adel, 1939b: Adel and Lampland, 1940a, 40b). These lines are weak between 10 μ and 11 μ, but much stronger near 8 μ or 14 μ. In regions between 8 μ and 13·5 μ not lying inside these lines, water vapour appears to produce a continuous absorption (Adel, 1939a), the transmitted radiation following an exponential law $e^{-k_\nu w}$. The absorption coefficient k_ν agrees with that predicted by Elsasser (§12, below) in the range 9–12 μ, but increases above it outside this. The measured absorptions, which are greater than Fowle's, are supported by residual-ray measurements made by Strong (1941).

Beyond 15 μ the variation of absorption with w has not been systematically studied. Weber and Randall (1932) measured the absorption between 16·5 μ and 25 μ of 3 m. of air, saturated at an unspecified temperature ($w \sim 0{\cdot}005$ cm. say). They find a mean absorption increasing from about 7 % between 17 μ and 19 μ to about 35 % near 25 μ. Hettner (1918) found the absorption at 81° c. of air containing about 0·003 cm. of water in the range 21–34 μ. His absorptions

* Dr. G. B. B. M. Sutherland has suggested to me that there is a possibility of N_2O absorption near 17 μ. *Added in Proof:* See, however, Adel (1943).

† The lines of CO_2 are close together, and, with few exceptions, fairly uniform in intensity; hence CO_2 absorption appears as continuous in the prismatic spectrum, save for powerful lines, like those at 12·65 μ and 13·3 μ, and even these appear as blends with water lines.

range from about 20 % near 21 μ to about 50 % near 34 μ. Strong and Woo (1932) found absorptions at various wave-lengths from 21 μ to 152 μ by residual-ray methods.

The absorptions measured in the experiments of Strong and Woo were found by comparing the transmissions through a cell of damp air containing about 0·0005 cm. of water and through an equal cell of dry air. The radiation in either case passed through a short length of atmospheric air, whose absorption cannot be ignored. Assuming that it contained about one-eighth as much water as the cell, the results indicate a maximum absorption of about 40 %, reached at about 100 μ.

§ 12. ELSASSER'S WORK

A systematic description of water-vapour absorption has been attempted by several workers. The most ambitious is that of Elsasser (1938a, 38b, 38d), based partly on computation and partly on observation. Elsasser approximates by taking the fractional absorption A in any wave-length range $\Delta\nu$ to follow the square-root law

$$A = \beta w^{\frac{1}{2}}$$

(w being measured in cm.) up to the point where this gives $A = 1$, and thereafter assumes total absorption. His aim is to determine β in any range.

For wave-lengths greater than 18 μ, Elsasser calculates A from the observed positions and computed intensities of absorption lines given by Randall, Dennison, Ginsburg and Weber (1937), using the formula of § 6:

$$A = \frac{2}{\Delta\nu}(\gamma w)^{\frac{1}{2}}\Sigma k^{\frac{1}{2}}.$$

The calculated values are subject to several sources of error: (i) Water-vapour lines tend to occur in close pairs or groups; the effect of overlapping in lines of a group cannot be neglected. (ii) The calculated line-intensities are approximate; they are obtained from the results for a symmetric rotator by the method described in § 4. The intensities of several lines are uncertain, the most uncertain being those arising from the transitions $J_\tau - (J-1)_\tau$, when $\tau \sim -\frac{1}{4}J$, and from the transitions $J_\tau - (J-1)_{\tau-4}$. The absorption of such lines, and of lines forbidden to a rigid symmetric rotator, is very important near 18 μ. (iii) The value adopted for the line width 2γ is 0·5 cm.$^{-1}$, Cornell's value for bands near 1 μ. We have already (§ 9) shown reason to doubt whether Cornell's values apply to bands at higher wave-lengths; and a systematic variation of γ with wave-length is not impossible. Because of all these sources of error, Elsasser's estimates cannot be expected to be correct in more than order of magnitude.

Elsasser found that in the range 18–33 μ

$$A = 10^{-106/\lambda + 4\cdot56} w^{\frac{1}{2}},$$

roughly, in the atmosphere, where λ is the wave-length in μ. This makes A small, even for radiation traversing the whole atmosphere, if $\lambda < 20\,\mu$. The observations of a black-out of solar radiation beyond 14 μ contradict this if, as is supposed, absorption beyond 17 μ is due to water-vapour. Such laboratory data as exist also disagree with Elsasser's result. At 21 μ, Strong (1941) gives $A = 3w^{\frac{1}{2}}$, and Weber and Randall (1932) and Hettner (1918) give concordant results;

Elsasser's formula gives $A = \frac{1}{2}w^{\frac{1}{2}}$ here.* Beyond 35 μ Elsasser's calculations give larger absorptions than are found from Strong and Woo's (1932) experiments ($\beta \sim 35$, say, as against $\beta \sim 15$).

Elsasser inferred from the Lorentz formula that the wings of strong lines beyond 35 μ should produce an appreciable continuous absorption between 8 μ and 22 μ. Adel actually observed an absorption roughly equal to that predicted, between 8·3 μ and 13·4 μ (§ 11). Strong (1941) showed that Adel's later measurements agree with the square-root law at least as well as with an exponential law, suggesting that the absorption may be, not a genuine continuous absorption, but a smoothed line absorption. Adel's grating observations (1941) do not suggest that water vapour gives a very large line absorption in the range, and the constancy of the observed absorption between 9 μ and 12 μ does not suggest line absorption. The numerical agreement between observation and prediction must, however, be regarded as fortuitous, since the Lorentz formula cannot be exact when $|\nu - \nu_0|$ is comparable with ν_0. The question whether the absorption is continuous or not must therefore be left open.

In the range 5–8 μ, Elsasser took the shape of the absorption band from Hettner's diagram (1918) and the central absorption from Fowle's work (1917).

Using his estimated absorptions, Elsasser has calculated the total emission of columns of damp air. His observed values agree roughly with these if $w < 0{\cdot}01$ cm., but are larger if $w > 0{\cdot}01$ cm., the discrepancy if $w > 0{\cdot}1$ cm. being about 25 % of the observed emission. His calculations of atmospheric emission are also below the observed values. For these reasons, Elsasser has greatly modified his results in a recent paper (1942), to which the reader is referred.

§ 13. SUMMARY OF OBSERVATIONS

To give an idea of the absorption at different wave-lengths, I have assumed that it is given, when greater than 10 %, by

$$A = \operatorname{erf}(\tfrac{1}{2}\beta(wm)^{\frac{1}{2}})$$

(roughly $A = \beta w^{\frac{1}{2}}$ if not too large), and have calculated β for room temperatures from the observations for different wave-length ranges. Roughly speaking, if $\beta < 1$ the atmosphere will always transmit a large part of the radiation of the corresponding wave-length; if $1 < \beta < 3$, the amount transmitted varies greatly with the humidity; if $\beta > 3$, there is very little transmission.

In the range 5–7·5 μ, v. Bahr's work (as interpreted by Hettner (1918)) is used. The value of w, not given by v. Bahr, is taken to be 0·002 cm.: this gives general agreement with Fowle's work (1917). Between 7·5 μ and 13·5 μ Adel and Lampland's results (1940 a, 40 b) are used. Between 13·5 μ and 17 μ we interpolate; between 17 μ and 25 μ, Weber and Randall's values (1932) are used, with $w = 0{\cdot}005$ cm. Between 21 μ and 34 μ, Hettner's results for 81° c. are used; to correct to room temperature we divide by a factor estimated as about 1·6 near 22 μ and 1·1 near 34 μ. We give results obtained from Strong and Woo's data (1932) separately. They give smaller values of β than Hettner for wave-lengths

* Elsasser's results seem to be affected by systematic errors of computation. In a recalculation I obtain results roughly agreeing with his beyond 28 μ, but ten times as large for 18–21 μ. Since between 18 μ and 100 μ the line strengths increase roughly in the ratio 500 : 1, the (square-root law) absorption should increase only in the ratio 22 : 1. Elsasser makes β increase from about 0·1 to 42, i.e. in the ratio 400 : 1.

between 21 μ and 34 μ, and smaller values than are given by Elsasser for larger wave-lengths: but it may be noted that residual-ray experiments made with steam suggest a transparency greater than their results show (Rubens and Hollnagel, 1910; Rubens and Wartenberg, 1911).

The values shown in table 2 do not claim great accuracy, both because of the approximate nature of the error-function formula and because of the difference between the results of different observers.

Table 2. Values of β

Range (μ)	β	Range (μ)	β	Range (μ)	β
5·0–5·2	1·7 (?)	7·8–8·0	0·9	25–27	4·6
5·2–5·4	2·7	8·0–8·2	0·5	27–30	6·5
5·4–5·6	4	8·2–8·4	0·3	30–34	9·0
5·6–5·8	7·5	8·4–8·6	0·25	(Strong and Woo)	
5·8–6·05	11	8·6–12·5	0·24	near 21 μ	2·7 (?)
6·05–6·2	8	12·5–13	0·3	27·3	4·5 (?)
6·2–6·35	3	13–14	0·4	32·8	5·5 (?)
6·35–6·45	8	14–15	0·5 (?)	41	11·5
6·45–6·7	11	15–16	0·7 (?)	52	15
6·7–6·9	7	16–17	1·0 (?)	63	14·5
6·9–7·2	4·5	17–19	1·5	80	11·5
7·2–7·4	3	19–21	2·1	94	16
7·4–7·6	2·2	21–23	3·5	120	15
7·6–7·8	1·4	23–25	4·0	152	13

§ 14. CONCLUSIONS

In theory we need experimental values of the absorption of water-vapour for all values of p, w and T. If the centres of absorption lines are blacked out, the number of independent variables can be reduced to two, since (§ 6) the absorption depends on p and w only in the combination pw.

A further reduction can be made if large absorptions are assumed to follow the error-function law (§ 6); i.e., if the fractional absorption in any small range of wave-lengths is given by a formula of the type

$$A = \mathrm{erf}\,(\tfrac{1}{2}\beta(pw\pi/p_0)^{\frac{1}{2}})$$

(p_0 = atmospheric pressure), or, when w is not too large, approximately

$$A = \beta(pw/p_0)^{\frac{1}{2}}.$$

In this case the whole run of the absorption is known, for a given temperature, if β is known. While this formula should give a reasonable approximation if the transmitted radiation is not too small, it is not likely to be exact. Further experiments on the variation of absorption with w and p are therefore desirable.

The variation with p has been very little studied. The variation with w in the 6·3 μ band has been determined fairly completely by Fowle, but further work with short columns of damp air might be helpful. In the rotation-band the absorption is not very well determined for any value of w, and here the need for further work is great. The difficulty of work in this range will be in part eased by the small resolution required: the mean absorption over a range of

wave-number 25 cm^{-1} is quite adequate. In the lower wave-lengths of the rotation band (below 19 μ, say), information regarding the joint absorption of water vapour and CO_2 would be valuable. In the range 8–13 μ, observations like Adel's seem most likely to give further information; but observations with columns of steam a few metres long may also help to decide whether there is a genuine continuous absorption in this range.

For meteorological work we need to know the absorption at temperatures below 0° C. Save for atmospheric observations of total emission, there are no experimental data regarding these temperatures, and, indeed, laboratory experiments are hardly practicable below 0° C. Extrapolation from temperatures above 0° C. to those below is possible, though risky: but the material for such extrapolation is at present lacking. Further calculations like those of Elsasser may be helpful, but for these further information regarding line breadths is required.

REFERENCES

Adel, A., 1939 a. *Astrophys. J.* **89**, 1.
Adel, A., 1939 b. *Astrophys. J.* **89**, 320.
Adel, A., 1941. *Astrophys. J.* **94**, 451.
Adel, A., 1943. *Astrophys. J.*, **96**, 239.
Adel, A. and Lampland, C. O., 1940 a. *Astrophys. J.* **91**, 1.
Adel, A. and Lampland, C. O., 1940 b. *Astrophys. J.* **91**, 481.
Adel, A., Slipher, V. M. and Barker, E. F., 1935. *Phys. Rev.* **47**, 580.
v. Bahr, E., 1909. *Ann. Phys., Lpz.*, (4), **55**, 476.
v. Bahr, E., 1913. *Verh. d. D. Phys. Ges.* **15**, 731.
Bernal, J. D. and Fowler, R. H., 1933. *J. Chem. Phys.* **1**, 515.
Brooks, F. A., 1941. *Mass. Inst. Tech. Papers in Phys. Oceanography and Met.* **8**, no. 2.
Cornell, S. D., 1937. *Phys. Rev.* **51**, 739.
Darling, B. T. and Dennison, D. M., 1940. *Phys. Rev.* **57**, 128.
Dennison, D. M., 1931. *Rev. Mod. Phys.* **8**, 22.
Elsasser, W. M., 1938 a. *Astrophys. J.* **87**, 496.
Elsasser, W. M., 1938 b. *Phys. Rev.* **53**, 768.
Elsasser, W. M., 1938 c. *Phys. Rev.* **54**, 126.
Elsasser, W. M., 1938 d. *Mon. Weath. Rev.* **66**, 175.
Elsasser, W. M., 1941. *Mon. Weath. Rev.* **69**, 1.
Elsasser, W. M., 1942. *Harvard Meteorological Studies*, No. 6.
Falckenberg, G., 1936. *Met. Z.* **53**, 172.
Falckenberg, G., 1938. *Met. Z.* **55**, 174.
Falckenberg, G., 1939. *Met. Z.* **56**, 72.
Fowle, F. E., 1917. *Smithson. Misc. Coll.* **68**, no. 8.
Hettner, G., 1918. *Ann. Phys., Lpz.*, (4), **55**, 476.
Hottel, H. C. and Mangelsdorf, H. G., 1935. *Trans. Amer. Inst. Chem. Engrs.* **31**, 517.
Kühne, J., 1933. *Z. Phys.* **84**, 72.
Ladenburg, R. and Reiche, F., 1913. *Ann. Phys., Lpz.*, (4), **42**, 181.
Margenau, H. and Watson, W. W., 1936. *Rev. Mod. Phys.* **8**, 22.
Mecke, R., 1933 *Z. Phys.* **81**, 313.
Nielsen, H. H., 1941. *Phys. Rev.* **59**, 565.
Randall, H. M., Dennison, D. M., Ginsberg, N. and Weber, L. R., 1937. *Phys. Rev.* **52**, 160.
Rubens, H. and Hollnagel, H., 1910. *Verh. dtsch. phys. Ges.* **12**, 83.
Rubens, H. and Wartenberg, H., 1911. *Verh. dtsch. phys. Ges.* **13**, 796.
Schmidt, E. and Eckert, E., 1936. *Forsch. a. d. Geb. d. Ing.-Wes.* **8**, 87.
Schnaidt, F., 1937. *Met. Z.* **54**, 234.
Schnaidt, F., 1939. *Gerl. Beit.* **54**, 203.
Strong, J., 1941. *J. Franklin Inst.* **232**, 1.
Strong, J. and Woo, S. C., 1932. *Phys. Rev.* **42**, 267.
Weber, L. R. and Randall, H. M., 1932. *Phys. Rev.* **40**, 835.
Wright, N. and Randall, H. M., 1933. *Phys. Rev.* **44**, 391.

THE SPECTRUM OF THE NIGHT SKY

By R. W. B. PEARSE,
Imperial College, London

§ 1. INTRODUCTION

On moonless nights the sky is found still to give a feeble illumination, which, for a horizontal plane, is about equivalent to that produced by a candle at a distance of 60 metres.

During the last thirty years strenuous efforts have been made to obtain and account for the spectrum of this radiation. In a brief report such as this there is insufficient space for justice to be done to the earlier work by which knowledge of the spectrum was increased step by step with improved spectrographs and with accumulating laboratory data. However, several summaries have already appeared: Déjardin (1936), Gauzit (1938), Nicolet (1939 a, 39 b) and Frongia (1939). These may be consulted for a description of this work and for a more complete bibliography than that given here.

Study of the lists of wavelengths, obtained by various observers, given in these summaries, shows that for a large proportion of the entries there is no agreement. Such entries are mostly those with low intensity values, and doubt arises whether they can be accepted as representing genuine features of the spectrum. Since the origin of some of the most prominent lines or bands is still obscure it appears wise to limit the discussion at this stage to those wavelengths which are reasonably certain and which are constant features of the spectrum. This is facilitated by the recent publication by Elvey, Swings and Linke (1941) of an excellent photograph of the night-sky spectrum obtained with the nebular spectrograph of the McDonald Observatory, together with a conservative list of wavelengths resulting from the measurement of several plates. Earlier photographs taken with the nebular spectrograph with a wider slit, which are useful for comparison, have been published by Struve, van Biesbroeck and Elvey (1937). In table 1 is given the list of wavelengths obtained by Elvey, Swings and Linke, together with possible assignments discussed in this *Report*.

§ 2. DESCRIPTION OF THE SPECTRUM

The majority of photographs of the night-sky spectrum show a background of continuum very similar to that due to scattered sunlight and in which the H and K absorption lines of Ca^+ may often be distinguished. The definition is probably not sufficiently good to show other Fraunhofer lines even if these are present in the radiation. At the long wavelength end of the spectrum there are four emission lines: λλ 5577, 5893, 6300 and 6364 A. The first is by far the strongest feature of the spectrum; the following two are usually the next in strength, although λ 5893 A. is very variable. These lines stand out well

above the continuum. In the remainder of the visible spectrum down to about λ 3800 A. there are a number of further radiations such as $\lambda\lambda$ 4420, 4174 and 4048 A., but these are nearer equality in intensity with the continuum. In the ultraviolet region the continuum is much less apparent, and there are a large number of further radiations which stand out clearly. Altogether Elvey, Swings and Linke have listed 52 radiations, of which 11 are somewhat less certain than the others.

§ 3. IDENTIFICATION OF RADIATIONS

The identification of these radiations presents some unusual difficulties. The feebleness of the source has so far permitted only the use of fast instruments of low dispersion and low resolving power. Consequently it is, in most cases, impossible to say from inspection of the spectrograms whether a given radiation is a line or a band, that is, whether it is atomic or molecular in origin. This difficulty is rendered more acute by the small extent of the individual bands, which is presumably due to the low temperature prevailing at the place of emission. Moreover, the lack of precision in the wavelength determinations makes it difficult to identify single lines or bands. From those proposed identifications which are fairly certain it appears that the spectrum is predominantly molecular in origin. The four radiations at the long wavelength end have, however, been proved to be due to atoms.

§ 4. ATOMIC LINES

The features of the night-sky spectrum identified with the closest approach to certainty are the lines $\lambda\lambda$ 5577, 6300, 6364 A., which are attributed to the forbidden transitions $^1D_2 - {}^1S_0$ and $^3P_{2,1} - {}^1D_2$ in the neutral oxygen atom O I, and the line λ 5893 A., which is a blend of the D lines of Na I, $^2S_{\frac{1}{2}} - {}^2P_{\frac{1}{2},\frac{3}{2}}$. The atomic nature of these lines has been established by the aid of the interferometer by Babcock (1923) and Bernard (1938). Other lines suggested by the presence of these, namely, λ 2972, $^3P_1 - {}^1S_0$ of O I and λ 3303 A. of Na I, have not been proved to be present. The oxygen line λ 2972 A. has the same initial state as the green line λ 5577 A., and should therefore be emitted according to the relative transition probabilities with a considerable fraction of its intensity. Since it is not observed it is presumably absorbed lower in the atmosphere. The line λ 3303 A. of Na is the second member of the series to which the D lines belong, and requires less additional energy than 2 ev. for its excitation. Its presence has not yet been conclusively established.

Kaplan (1938) has obtained a line at λ 3471 A. in an afterglow of nitrogen which he ascribes to the forbidden transition, $^4S - {}^2P$, of N I, and proposes this as a possible identification of one of the radiations of the night sky. Gauzit has supported this, but Elvey, Swings and Linke find the value λ 3460 A. for the wavelength of the n.-s.* band, which is too great a departure from the value observed by Kaplan, and also from the calculated value for the transition given by Bowen (1936), for the identification to be correct. Further evidence on this point could be obtained by a search for the line λ 10407 A., which has the same initial ^2P level.

* It is convenient to use the abbreviation "n.s." for "night sky" when discussing individual bands.

Several other suggestions have been put forward. Thus it was formerly claimed that lines of the rare gases were present, and there are a number of close coincidences of wavelengths, but to these there is the objection that they require far greater energy for their excitation than appears to be available, judging from the rest of the spectrum.

§ 5. MOLECULAR BANDS

Apart from the atomic lines already discussed, it appears likely that most of the remaining radiations are due to molecules. It has, however, up to the present, proved unusually difficult to identify them with a satisfactory degree of certainty. For this the low dispersion and resolving power are chiefly responsible, but, in addition, there is the uncertainty arising from the lack of knowledge of the process of excitation, which may give an unusual vibrational intensity distribution. In the following discussion the more plausible suggestions are briefly considered.

Oxygen. O_2. The emission bands of the Schumann-Runge system, $B\,^3\Sigma \rightarrow X\,^3\Sigma$, of O_2 have been analysed by Lochte-Holtgreven and Dieke (1929), who observed bands arising from the $v'=0$, 1 and 2 levels. Comparison of these with the n.-s. bands shows numerous coincidences. Further coincidences are shown with calculated wavelengths for the $v'=3$ level. It is also worth considering values calculated for $v'=-1$ and -2, since, with band systems involving large changes of internuclear distance, the values of the vibrational quantum numbers are rather uncertain in the absence of observation of the vibrational isotope effect. When this is done, some further coincidences are obtained which fit in with the general run of the Condon parabola. It is desirable that an attempt should be made to observe this extension of the system in the laboratory for direct comparison with the night-sky spectrum. The measure of the agreement of the wavelengths can be seen from table 1. The system provides the most probable explanation of a large number of the bands.

The $A\,^1\Sigma \rightarrow X\,^3\Sigma$ system of O_2 which appears in absorption in the solar spectrum giving the well-known bands at $\lambda\lambda$ 7594 and 6867 A. does not appear to contribute to the night-sky emission. The suggestion made by Elvey, Swings and Linke (1941) that the (3, 0) band with head at λ 5788 A. and origin at λ 5796 A. may explain the n.-s. bands $\lambda\lambda$ 5775 and 5816 A. observed by Babcock (1939) appears very doubtful.

Ozone. O_3. Laboratory data on the emission of ozone are very meagre. This is unfortunate, since it is not safe to assume that the emission spectrum of a polyatomic molecule will bear a close resemblance to the absorption spectrum. Bands attributed to ozone have been observed in emission by Janin (1938) and Johnson (1924). The lists of wavelengths given by these authors do not agree and, therefore, may in part at least have different origins; but as the wavelengths were obtained in association with oxygen and ozone, and as some of them approximate to the night-sky values, the spectra appear worthy of further investigation.

Nitrogen. N_2. The Vegard-Kaplan system, $A\,^3\Sigma \rightarrow X\,^1\Sigma$, accounts for several of the stronger bands. The long wavelength end of the system was observed in the laboratory by Bernard (1935), who excited a mixture of nitrogen and

argon with an electron beam. The shorter wavelengths had be n observed previously by Kaplan (1934). These bands are included in table 1.

The First Positive system, $B^3\Pi \rightarrow A^3\Sigma$, has been held to account for some of the bands at the red end of the spectrum (Elvey, Swings and Linke, 1941; Déjardin, 1936), but the evidence in support of this does not appear very strong.

With regard to the presence of the Second Positive system, $C^3\Pi \rightarrow B^3\Pi$, in the night-sky spectrum, there has been considerable difference of opinion (Gauzit, 1938). Gauzit concludes that it is present. The strongest band of the system under usual conditions is the (0, 0) band at λ 3371 A. degraded to the violet, and this has been held to contribute, at least in part, to the n.-s. band near that wavelength. Elvey, Swings and Linke find for this band, however, the wavelength λ 3378 A., and point out that it is sharp at the violet edge. They conclude, therefore, that the system, if present, is of extreme weakness. It must be remembered, however, that this system, as well as the First Positive system, has been definitely observed in the spectrum of the aurora, and these may, therefore, appear on plates exposed to the night sky during a weak unsuspected auroral display. The method of Slipher (1933) of taking simultaneous exposures at the zenith and above the N., S., E. and W. horizons has the advantage of showing up the auroral bands by the much greater variation in intensity in the different directions.

The Lyman-Birge-Hopfield system, $a^1\Pi \rightarrow X^1\Sigma$, has been observed both in absorption and emission in the vacuum ultraviolet region, occurring readily in discharges. With high current density discharges the system has been extended to about λ 2500 A. Elvey, Swings and Linke have calculated the extension of the system to longer wavelengths, obtaining some coincidences with n.-s. bands, notably $\lambda\lambda$ 3378 and 3834 A., which are degraded to the red, as would be required for this system. It is difficult, however, to see why the long wavelength end of this system should be strong in the night-sky spectrum when it does not ordinarily accompany the system when strongly developed in the laboratory. An attempt to produce the bands would be of interest, as suggested by these authors. Some of the other band systems reported (see Pearse and Gaydon, 1941) in connection with nitrogen might also repay further investigation.

N_2^+. The Negative system of nitrogen, $B^2\Sigma \rightarrow X^2\Sigma$, is strongly developed in the spectrum of the aurora. It has also been observed in the spectrum of the illuminated upper atmosphere at dawn and dusk (Slipher, 1933). When light from these sources is absent from the night-sky spectrum, the system does not appear, or at least becomes very weak. The (0, 0) band at λ 3914 A. is not a conclusive test, since the (1, 18) band of the Schumann-Runge system falls at λ 3913 A., but Elvey, Swings and Linke show no entries at the positions of the (1, 0) and (0, 1) bands, $\lambda\lambda$ 3582 and 4278 A.

Nitric oxide. NO. The bands of the γ system, $A^2\Sigma \rightarrow X^2\Pi$, of NO are degraded to the violet. The system is readily obtained in a discharge tube, but, unfortunately, all the strongest bands are too far to the ultraviolet to provide a sensitive test for NO in the night sky. The strongest band available is the double-headed band (0, 6) $\lambda\lambda$ 3009 and 2997 A. just outside the range observed by Elvey, Swings and Linke. The correcting plate of the Schmidt camera which

they used was of ultraviolet glass which begins to absorb around λ 3200 A., and so weakens the region to shorter wavelengths.

The double-headed bands of the β system, $B^3\Pi \rightarrow X^3\Pi$, are degraded to the red and appear strongly in the afterglow of active nitrogen containing a small quantity of oxygen.

The degree to which n.-s. bands approximate to observed bands in the two systems of NO may be seen from table 1.

Table 1. Spectrum of the night sky

Night sky		O_2 Schumann-Runge system			N_2 Vegard-Kaplan system			NO γ system		NO β system		
λ (A.)	I	v′, v″	λ (A.)	I	v′, v″	λ (A.)	I	v′, v″	λ (A.)	v′, v″	λ (A.)	I
					0, 7	2936	9					
					1, 8	2997	2					
					5, 11	3068						
(3110)		0, 12	3104	4				1, 8	3112			
					3, 10	3128		3, 11	3121			
					6, 12	3137						
3145	1											
(3157)		1, 13	3162									
					1, 9	3197	5	0, 7	3171			
3211	3				7, 13	3209		2, 10	3201	0, 8	3198 3207	10
3233		0, 13	3233	4								
3263	1				5, 12	3270		1, 9	3278			
3298	3	1, 14	3294					3, 12	3303	4, 11	3290 3299	0
		−1, 13	3309									
3321	2									2, 10	3331 3340	3
					6, 13	3345						
		2, 15	3357	3	0, 9	3353						
3378	5	0, 14	3370	5				0, 8	3375	0, 9	3376 3386	10
		−2, 13	3391									
					4, 12	3417				3, 11	3399 3408	1
3425	1	1, 15	3433		1, 10	3425	4					
3460	2	−1, 14	3453					1, 10	3458	1, 10	3446 3457	2
3488	3											
		2, 16	3500	2	2, 11	3503	3					
		0, 15	3516	5								
		−2, 14	3543									
3556•10												
		1, 16	3583	3	3, 12	3582	2			0, 10	3572 3583	10
(3598)												
		−1, 15	3607		0, 10	3603	4					
3623	1											
3636	2											
		2, 17	3651									
3664	1				4, 13	3664	1			1, 11	3647 3658	4
		0, 16	3673	5	1, 11	3683	2					
3707	3	−2, 15	3705									
3742	4	1, 17	3742	5								
					5, 14	3750	1					
		−1, 16	3772		2, 12	3768	1					
3787	1									0, 11	3788 3801	10
3817	1	2, 18	3813									

Table 1 (*cont.*). Spectrum of the night sky

Night sky		O_2 Schumann-Runge system			N_2 Vegard-Kaplan system			NO γ system		NO β system		
λ (A.)	I	v′, v″	λ (A.)	I	v′, v″	λ (A.)	I	v′, v″	λ (A.)	v′, v″	λ (A.)	I
3834	4	0, 17	3841	6								
		−2, 16	3879		3, 13	3855	3			1, 12	3868	5
		3, 19	3887		0, 11	3889	4				3881	
3912	5	1, 18	3913	5								
					7, 16	3940	4					
3948	3	−1, 17	3950		4, 14	3950				2, 13	3950	2
3984	2	2, 19	3987	4	1, 12	3979	3				3963	
4016	2	0, 18	4021	5								
4048	2				5, 15	4045				0, 12	4027	6
		3, 20	4064								4042	
4072	5	−2, 17	4067		2, 13	4072	3					
		1, 19	4095	6								
(4107)										1, 13	4114	4
4134	3	−1, 18	4140		6, 16	4144	2				4128	
4174	5	2, 20	4173	6	3, 14	4171	4					
		0, 19	4214	5	0, 12	4219	1			2, 14	4201	4
4259	1	3, 21	4253		7, 17	4253					4215	
(4268)		−2, 18	4269		4, 15	4274						
		1, 20	4292	5						3, 15	4288	3
											4303	
4316	2				1, 13	4320	2			0, 13	4294	
		−1, 19	4345								4310	3
4379	1	2, 21	4372	5						1, 14	4386	
4420	5	0, 20	4423		2, 14	4425					4401	2
		3, 22	4455									
		−2, 19	4488		6, 17	4495	1			2, 15	4480	
4543	3				3, 15	4535	5				4496	3
		−1, 20	4567									
4587	1	2, 22	4587							3, 16	4574	
					0, 13	4605	2				4591	5
					7, 18	4616	5					
					4, 16	4650	4					
4669	1	3, 23	4672									
		2, 23	4816		1, 14	4718	2			2, 16	4791	
4827	3				2, 15	4837	8				4840	1
4878	1											
					3, 16	4960	2			3, 17	4892	
(4994)											4912	2
					0, 14	5060	4					
5213												
					2, 16	5326	0			3, 18	5253	
(5352)											5270	0
(5441)												
5577	500											
(5750)												
5893	25											

* On one plate this line appeared to be separated into two lines with the wavelengths 3549 A. and 3558 A.

N.B.—The wavelengths in brackets in the night-sky spectrum are uncertain. The wavelengths without intensities in the spectra of O_2 and N_2 are calculated values. Those in the spectrum of NO have been observed in the laboratory but are very weak.

Other oxides of nitrogen. Little is known regarding the emission spectra of other oxides of nitrogen, and so far no evidence has been presented to show that they add anything to the night-sky spectrum. Their absorption spectra are discussed by Price in another section of these *Reports*.

Hydrogen and hydrides. There appears in the night-sky spectrum no radiation ascribed either to the H atom or the H_2 molecule. On the other hand, it has been suggested that some of the radiations might be due to such hydride molecules as CH, NH, OH and OH^+. Since for hydride band systems the (0, 0) band is often the only one to appear strongly, these systems are difficult to establish as present unless some of the band structure is resolved. The hydrides are therefore best considered together.

For CH three band systems are known: $A\,^2\Delta \rightarrow X\,^2\Pi$, (0, 0) λ 4312 A.; $B\,^2\Sigma \rightarrow X\,^2\Pi$, (0, 0) λ 3889 A.; and $C\,^2\Sigma \rightarrow X\,^2\Pi$, (0, 0) λ 3144 A., (1, 1) λ 3157 A. Of these the third is ordinarily much weaker than the other two. Elvey, Swings and Linke consider it probable that the n.-s. bands at λλ 4316, 3157 and 3145 A. are accounted for by the first and last of these systems, the second system being absent. No reason is put forward, however, to account for this unusual intensity distribution.

The system of NH which most readily appears is the $^3\Pi \rightarrow {}^3\Sigma$ system (0, 0) λ 3360 A., (1, 1) λ 3371 A. Of these two bands, λ 3360 A. is considerably the stronger. It therefore appears unlikely that these can be considered to contribute to the n.-s. band λ 3378 A., which is sharp on the violet side.

The system of OH, $^2\Sigma \rightarrow {}^2\Pi$, (0, 0) λ 3064–3089 A. comes in the region where the intensity of the night-sky spectrum is rapidly diminishing; there is no definite evidence of its presence.

The molecule OH^+ is known to possess a system, $^3\Pi \rightarrow {}^3\Sigma$, (0, 0) λ 3565 A. analogous to that of NH. This band is in the region of the strong n.-s. band λ 3556 A., but since the OH^+ band is degraded to the red the wavelength discrepancy appears to be too large for them to be identical.

Carbon compounds. Kaplan (1937) and Cabannes, Dufay and Gauzit (1938) have concluded that CN bands are present in the night-sky spectrum on account of coincidences between n.-s. bands and the tail bands of the $B\,^2\Sigma \rightarrow X\,^2\Sigma$ system. Elvey, Swings and Linke have concluded that they are absent on account of the absence of the three sequences headed by the (1, 0), (0, 0) and (0, 1) bands at λλ 3590, 3883 and 4216 A. The latter conclusion seems preferable. Stronger proof than the wavelength coincidences put forward is necessary before such an unusual intensity distribution can be accepted.

Similarly the Ångström system of CO, $B\,^2\Sigma \rightarrow A\,^2\Pi$, may be considered absent as the (0, 0) λ 4511 A., (0, 1) λ 4835 A., (0, 2) λ 5198 A. and (0, 3) λ 5610 A. bands do not appear. The Swan system of C_2, $^3\Pi \rightarrow {}^3\Pi$, in the absence of the bands (1, 0) λ 4737 A., (0, 0) λ 5165 A., (0, 1) λ 5635 A., may also be assumed to be absent.

The molecule CO_2 has a complex emission spectrum, which has not been analysed, giving numerous bands in the region from λ 3900 A. to λ 5500 A. The spectrum is obtained in the flame of carbon monoxide burning in air or oxygen and in the afterglow of a heavy current discharge through carbon dioxide. There is no evidence for its presence in the night-sky spectrum.

Ionized molecules. Besides the molecules N_2^+ and OH^+, which have already been considered, there are three more, namely, O_2^+, CO^+ and CO_2^+, which possess band systems within the range of spectrum under discussion.

For O_2^+ there are two systems: the First Negative system, $b\,^4\Sigma \rightarrow a^4\Pi$, degraded to the violet with strongest bands (0, 1) λ 6419 A., (0, 0) λ 6026 A., (1, 0) λ 5632 A. and (2, 0) λ 5296 A.; and the Second Negative system, $A\,^2\Pi \rightarrow X\,^2\Pi$, degraded to the red with a large number of bands covering the whole region from λ 6100 A. to λ 2000 A. Nothing in the night-sky spectrum appears to correspond to the bands of the First Negative system, which has an excited lower electronic state, but the following double bands of the Second Negative system are close to n.-s. bands: (2, 5) λλ 3211, 3231 A.; (1, 5) λλ 3300, 3322 A.; (2, 7) λλ 3594, 3620 A.; (1, 7) λλ 3707, 3734 A. The lower level of this system is the lowest state of the ion.

For CO^+ there are also two systems in this range, the Comet-tail system, $A\,^2\Pi \rightarrow X\,^2\Sigma$, and the Baldet-Johnson system, $B\,^2\Sigma \rightarrow A\,^2\Pi$. The former system is degraded to the red and is based on the lowest state of the ion. The bands are double-headed, with the (0, 0) band-heads at λλ 6405, 6354 A. The latter system is degraded to the violet and is based on the initial state of the Comet-tail system. The bands are double double-headed, with the (0, 0) band-heads at λλ 3954, 3957, 3973 and 3978 A. These systems of CO^+ have neither been obtained in the night-sky spectrum nor in the spectrum of the aurora.

For CO_2^+ there is an extensive system of bands, formerly attributed to CO_2, which appears readily in oxygen discharge tubes containing a trace of carbon impurity. Several groups of narrow bands degraded to the red appear in the region of the spectrum from λ 2800 A. to λ 5000 A. The system certainly does not appear in its entirety in the night-sky spectrum, but it is noteworthy that it contains strong bands at λ 3377 A. and at λλ 3546, 3551 and 3562 A., which are near the outstanding bands λ 3378 A. and λ 3556 A. of the night sky.

Thus at present the spectrum provides no conclusive evidence for the presence of ions in the night sky as a normal feature. The bands of CO^+ and OH^+ appear to be definitely absent, those of N_2^+ are absent or very weak except during twilight or auroral displays, while the coincidences in wavelengths between bands of the night sky and those of O_2^+ and CO_2^+ may well be due to chance. Of course it may be argued that a considerably greater amount of energy is required to produce these band systems than to produce the Vegard-Kaplan and Schumann-Runge systems of N_2 and O_2. But such arguments involve the tacit assumption that the whole of the night-sky spectrum has the same place of origin.

Unidentified bands. It is a rather striking fact that λ 3556 A., the strongest band in the ultraviolet region of the spectrum, has not been identified. Moreover, the band λ 3378 A. appears to be rather too intense to be wholly explained by the proposed identifications. In addition to these, Elvey, Swings and Linke list λλ 3263 (1), 3321 (2), 3817, 4587 (1), 4669 (1), 4878 (1) and (4994) A. without identification. It is possible that three of these, λλ 3817, 4587 and 4669 A., belong to the Schumann-Runge system, as shown in table 1, but the others remain unidentified. The presence of these strong unidentified bands in the night-sky spectrum has led Gauzit and Cabannes to suggest that they belong to a new

system of bands of nitrogen (Gauzit, 1938). This is obviously a point for further laboratory investigation.*

§6. ABSORPTION

A question concerning the night-sky spectrum which seems to require further investigation is to what extent absorption plays a part in its production. If, as appears certain, the emission spectrum is derived from the upper atmosphere, then the radiation has to pass through the ozone layer and the oxygen of the lower atmosphere. Consequently the absorption bands of ozone and oxygen must be superimposed on the emission bands. The same is also true of any other absorbing constituent of the atmosphere. This may provide an explanation of a rather unsatisfactory feature in identification of the Vegard-Kaplan and Schumann-Runge systems, which is that some bands of these systems appear to be absent from the night-sky spectrum, although other bands with the same initial level are present, and although both have similar intensities in the laboratory. In a few cases such absences occur at the same wavelength region for both systems. Thus the Schumann-Runge bands (2, 16) λ 3500 A. (2) and (0, 15) λ 3516 A. (5) and the Vegard-Kaplan band (2, 11) λ 3503 A. (3) are absent, although the Schumann-Runge band (0, 13) λ 3233 A. (4) and the Vegard-Kaplan band (2, 13) λ 4072 A. (3) are present.

The observation of the bands of N_2^+ at twilight shows that this ion is produced in the sunlit atmosphere, and raises the question of the possibility of observing it in absorption in the solar spectrum. The strongest absorption may be expected at λ 3914 A.. with weaker absorption at λλ 3582 and 3308 A.

§7. FUTURE INVESTIGATION

For satisfactory identification of the bands of the night-sky spectrum it is desirable that the spectrograms should be improved until some at least of the structure of the bands is resolved. Since the nebular spectrograph is able to photograph the night-sky spectrum in a few hours, there seems to be every reason for supposing that this could be done, since some speed could, if necessary, be sacrificed for higher dispersion. Moreover, it would be interesting to make observations on the variation of intensity of selected members of the band systems, throughout the night and throughout the year, such as have been made for the oxygen and sodium lines.

Further laboratory investigations of the spectra of O_2 and N_2 are desirable, and attempts should be made to obtain the emission spectra of ozone and the oxides of nitrogen. A search along these lines may yield a satisfactory identification of the strong ultraviolet band at λ 3556 A.

* Since the above was written, P. Swings (1943), following a suggestion of J. Dufay, has made a tentative assignment of the majority of the ultraviolet bands, including λ 3556, to the forbidden $^3\Sigma_u^+ \rightarrow {}^3\Sigma_g^-$ system of O_2. So far this system has only been known from absorption bands in the region λ 2400–λ 2900, obtained with long columns of oxygen, but possible emission bands in the longer wavelength region have been calculated by extrapolation. The suggestion is attractive, but the table of assignments is not convincing.

REFERENCES

BABCOCK, H. D., 1923. *Astrophys. J.* **57**, 209.
BABCOCK, H. W., 1939. *Pub. A.S.P.* **51**, 47.
BERNARD, R., 1935. *C.R. Acad. Sci., Paris*, **200**, 2074.
BERNARD, R., 1938. *C.R. Acad. Sci., Paris*, **206**, 928.
BOWEN, I. S., 1936. *Rev. Mod. Phys.* **8**, 55.
CABANNES, J., DUFAY, J. and GAUZIT, J., 1938. *Nature, Lond.*, **142**, 755.
DÉJARDIN, G., 1936. *Rev. Mod. Phys.* **8**, 1.
ELVEY, C. T., SWINGS, P. and LINKE, W., 1941. *Astrophys. J.* **93**, 337.
FRONGIA, G., 1939. *Nuovo Cim.* **16**, 360.
GAUZIT, J., 1938. *Ann. Astrophys.* **1**, 334.
JANIN, J., 1938. *C.R. Acad. Sci., Paris*, **207**, 145.
JOHNSON, R. C., 1924. *Proc. Roy. Soc.* A, **105**, 683.
KAPLAN, J., 1934. *Phys. Rev.* **45**, 675.
KAPLAN, J., 1937. *Nature, Lond.*, **52**, 1552.
KAPLAN, J., 1938. *Nature, Lond.*, **141**, 645 and 1139.
LOCHTE-HOLTGREVEN, W. and DIEKE, G. H., 1929. *Ann. Phys., Lpz.*, **3**, 937.
NICOLET, M., 1939 a. *Mém. Inst. Mét. Belg.* **9**.
NICOLET, M., 1939 b. *Bull. Acad. Belg. Cl. Sci.* **25**, 81.
PEARSE, R. W. B. and GAYDON, A. G., 1941. *The Identification of Molecular Spectra* (London : Chapman and Hall).
SLIPHER, V. M., 1933. *Mon. Not. R. Astr. Soc.* **93**, 657.
STRUVE, O., BIESBROECK, G. VAN and ELVEY, C. T., 1938. *Astrophys. J.* **87**, 559.
SWINGS, P., 1943. *Astrophys. J.* **97**, 72.

IONIZATION PHENOMENA IN THE EARTH'S ATMOSPHERE

By J. SAYERS,
The University, Birmingham

§1. RECOMBINATION AND THE IONIZATION BALANCE IN THE LOWER ATMOSPHERE

In the lower atmosphere, for the first 20 km. above sea-level, the study of the ionization balance is simple. This is due to three factors: (i) the source of ionization, assumed to be solely cosmic rays in this region, does not vary appreciably with time; (ii) the electrons formed live for such a short time that the part played by them in recombination or other processes may be ignored, or alternatively, we may say that the probability of attachment per electron is very much greater than the probability of recombination; (iii) the concentration gradient is everywhere so small that diffusion and drift of the ions in the electric field are of negligible importance, and ionic recombination alone is operative as the destructive process.

To justify (ii), an approximate calculation, using Cravath's (1929) or Bradbury's (1933) figures for the rate of electron attachment in air, shows that the mean life of free electrons varies from 10^{-5} or 10^{-6} sec., near ground-level, to about 10^{-4} sec., at 18 km.; whereas the corresponding lifetimes of negative ions, before being lost by recombination, are 250 and 200 sec. For (iii) it is sufficient to observe that an ion will, on the average, move only a few metres under the influence of the combined forces of diffusion and the electrostatic field, during its lifetime; and in this distance there is no appreciable change in the ionization density.

With the above assumptions the equation

$$N_0 = (q/\alpha)^{1/2} \qquad \ldots\ldots(1)$$

defines the relation between N_0, the equilibrium number of pairs of ions per c.c.; q, the rate of ion production per c.c. per sec.; and α, the recombination coefficient. If we wish to calculate the electrical conductivity in any region of the atmosphere we also require to know the mobility u, and for this calculation we assume that u varies directly as the absolute temperature T and inversely as the pressure p. Although the temperature dependence of u is more complicated than this simple law, the error introduced will be small within the restricted range of temperature under consideration. We therefore have

$$u = u_0 \frac{T}{273} \frac{760}{p} \qquad \ldots\ldots(2)$$

and the conductivity

$$k = eu_0 \frac{T}{273} \frac{760}{p} N_0, \qquad \ldots\ldots(3)$$

where u_0 is the ionic mobility at S.T.P. But

$$N_0 = (q/\alpha)^{1/2}$$
$$= \left\{ \frac{q}{\alpha_{273}} \frac{p}{760} \left(\frac{T}{273} \right)^2 \right\}^{1/2}, \qquad \ldots\ldots(4)$$

where q_0 is the value of q reduced to S.T.P., and α_{273} is the recombination coefficient for the particular value of p, but corrected to 0° c. The complete expression for conductivity is, therefore,

$$k = 1{\cdot}50 \times 10^{-3} \frac{eu_0 T^{7/4}}{p^{1/2}} \left(\frac{q_0}{\alpha_{273}} \right)^{1/2} \qquad \ldots\ldots(5)$$

Experimental measurements of the conductivity k have been made by Gish and Sherman (1936). They constructed an automatic recording apparatus which was taken aloft on the stratosphere flight of "Explorer II" in South Dakota on 11th November, 1935, by Stevens and Anderson (1938). For comparison with their direct measurements we may calculate values of k by equation (5) at various heights in the atmosphere. We employ Regener's measurements of q obtained with sounding balloons at approximately the same latitude as Dakota, and for α we use recent determinations by Sayers (1938). Figure 1 shows the relation between the calculated and observed conductivity

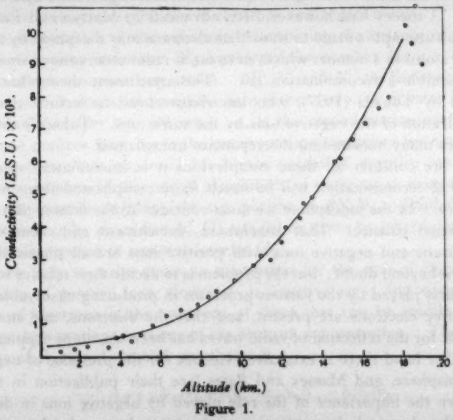


Figure 1.

at various heights. In both cases the graphs are for the conductivity of the positive ions, as for these the observations of the stratosphere flight are the more complete. In the region between 15 and 17 km., where the agreement is not very exact, the observed values were not directly recorded, but derived from the negative conductivity; so there is the possibility of a slight error in this region. However, the general agreement indicates the exactness of our knowledge of these fundamental electrical processes occurring in the lower atmosphere.

§ 2. RECOMBINATION IN THE IONOSPHERE

(i) *Theoretical considerations*

At altitudes very much greater than it is possible to reach with sounding balloons there exist regions of intense ionization responsible for the reflection of radio waves. Of these, the most important are the E and F regions. Each region may, however, assume considerable complexity; for example, during the day the E region contains not only the normal E layer, situated at a height of about 110 km., which behaves in a comparatively regular manner, but also an ionization sometimes known as "sporadic E" or "intense E", which may fluctuate greatly in intensity. The F region during the day may be composed of two layers or, rather, locations of large ionization gradient, the F_1 and F_2 layers. The F_1 layer is usually at a height of about 200 km., and the F_2 layer at 230 to 300 km. These two components of the F region often merge (during winter, for example) into a substantially single layer.

The problems of the agencies responsible for the production of the ionized layers, the mechanism of disappearance of electrons, and the ion content of the ionosphere have all been the subject of many investigations and theories. Appleton and Naismith (1935) showed that the normal E-region ionization is produced by the ultra-violet light from the sun. and is governed by a recombination law. A theory has, however, been advanced by Martyn and Pulley (1936) in which an attempt is made to show that electrons may disappear by attachment to neutral atoms in a manner which, in so far as radio observations are concerned, would resemble a recombination law. The attachment theory has also been advocated by Tukada (1937), who has endeavoured to include the effects of photo-ionization of the negative ions by the sun's rays. Tukada's theory, however, leaves many outstanding discrepancies unexplained.

When we consider all these complexities it is immediately apparent that the study of recombination will be much more complicated than in the lower atmosphere. In the ionosphere we must consider all the factors that determine the ionization balance. That attachment, detachment and recombination of both electrons and negative ions with positive ions are all physically possible processes is beyond doubt; but the problem is to decide their relative importance, and the parts played by the various processes in producing observable effects.

That free electrons are present, and that the electrons, and not ions, are responsible for the reflection of radio waves has been shown by Appleton (1933). On the other hand there is extensive evidence for the presence of negative ions in the ionisphere, and Massey and Bates (see their publication in this series) have shown the importance of the role played by negative ions in determining many of the electrical characteristics of the ionosphere. Appleton (1937) has pointed out that negative ions, if they exist in sufficient concentration, would provide an appreciable contribution to the direct-current conductivity of the ionosphere, and would therefore contribute to the observed variations in terrestial magnetism.

Immediately we recognize these facts, we are confronted with the task of explaining why, in the ionosphere, all the electrons do not quickly disappear by attachment to oxygen atoms which undoubtedly constitute a considerable

proportion of the neutral particles. The first, and possibly the most obvious, suggestion is that since the attachment energy of electrons to oxygen atoms is about 2 or 3 volts, there will be an abundance of quanta in the sun's rays capable of re-ionizing the negative ions. This explanation, however, still leaves the persistence of free electrons during the night unexplained, since the rapid attachment of electrons would then be uncompensated by photo-detachment.

The first step towards the elucidation of these difficulties was the suggestion by Martyn and Pulley (1936) of a process of electron detachment which would be effective by night as well as by day. The process proposed by them was the chemical combination of an oxygen negative ion and an oxygen atom, with the liberation of a free electron and the formation of a neutral oxygen molecule, viz.,

$$O^- + O \rightarrow O_2 + e. \qquad \ldots\ldots(6)$$

Other processes of this type are

$$O^- + N \rightarrow NO + e \qquad \ldots\ldots(7)$$

and

$$O^- + N_2 \rightarrow N_2O + e. \qquad \ldots\ldots(8)$$

These detachment reactions have been investigated in greater detail by Massey (1937, 38), who has shown that the processes of electron attachment and detachment are so effective that they will determine a quasi-stationary balance between the concentrations of electrons and negative ions.

For the formulation of the equations determining the balance of ionization we use the following notation: N_0 represents the number per c.c. of neutral particles to which electrons may attach. N_+, N_e and N_- are the concentrations of positive ions, electrons, and negative ions respectively. q denotes the rate of production of ions (per c.c. per sec.). α_e is the recombination coefficient for electrons and positive ions. α_i is the corresponding coefficient for negative ions and positive ions. λ is the ratio of concentrations of negative ions to electrons. β is a measure of the rate of attachment of electrons to neutral particles such that the number attaching per c.c. per sec. is $\beta N_e N_0$. Similarly, let the rate of detachment by processes (6), (7) and (8) be given by $\gamma N_- N_0$, and let the rate of photo-ionization of negative ions be IN_-.

It is now easy to write down an expression for the rate of change of the concentration of positive ions, electrons, or negative ions; but since the concentration of electrons is the only one of these that can be readily observed, we develop the equations in terms of the electron concentration. The equation for the rate of change of electron population is

$$\frac{dN_e}{dt} = q + \gamma N_- N_0 + IN_- - \alpha_e N_e N_+ - \beta N_e N_0. \qquad \ldots\ldots(9)$$

Since, as shown by Massey, $\beta N_e N_0$ and $(N_- N_0 \gamma + IN_-)$ are each much larger than any of the other terms, the ratio of these is approximately equal to unity; therefore

$$\lambda = \frac{N_-}{N_e} = \frac{\beta N_0}{\gamma N_0 + I}, \qquad \ldots\ldots(10)$$

and, assuming the medium to be electrically neutral, we may write

$$N_+ = (1 + \lambda) N_e. \qquad \ldots\ldots(11)$$

The balance equation for negative ions is

$$\frac{dN_-}{dt} = \beta N_e N_0 - \gamma N_- N_0 - IN_- - \alpha_i N_- N_+. \qquad \ldots\ldots(12)$$

Combining this with equation (9), we get

$$\frac{dN_e}{dt} = q - \alpha_e N_e N_+ - \alpha_i N_- N_+ - \frac{dN_-}{dt}. \qquad \ldots\ldots(13)$$

Using relations (10) and (11) this reduces to

$$\frac{dN_e}{dt} = \frac{q}{1+\lambda} - (\alpha_e + \lambda\alpha_i)N_e^2 - \frac{N_e}{1+\lambda}\frac{d\lambda}{dt}. \qquad \ldots\ldots(14)$$

In the absence of an ionizing process, and grouping all the quantities under N_e^2, we get

$$\frac{dN_e}{dt} = -N_e^2\left\{\alpha_e + \lambda\alpha_i + \frac{1}{N_e}\frac{d}{dt}\log(1+\lambda)\right\}; \qquad \ldots\ldots(15)$$

the factor multiplying N_e^2 we call the "effective recombination coefficient" α'.

It will be readily appreciated that unless λ is changing rather rapidly with time the term

$$\frac{1}{N_e}\frac{d}{dt}\log(1+\lambda)$$

will be small, and we may write for the effective recombination coefficient,

$$\alpha' = \alpha_e + \lambda\alpha_i \qquad \ldots\ldots(16)$$

(an expression given in recent papers by Appleton and Sayers (1938), and by Massey and his collaborators (1939)).

This will represent the conditions, except possibly at sunrise and sunset. when the term I in λ, involving the photo-ionization of negative ions, will be changing rapidly.

We deduce from such considerations that the presence of negative ions adds a term $\lambda\alpha_i$ to the apparent electron recombination coefficient, and in addition (see equation (14)), it reduces the apparent rate of ion production, measured in terms of electron concentration, in the ratio $1 : (1+\lambda)$.

(ii) *Recombination in region "E": results of radio observations*

Under equilibrium conditions equation (14) above becomes

$$\frac{q}{1+\lambda} = \alpha' N_e^2. \qquad \ldots\ldots(17)$$

Now according to the theory of Chapman (1931) relating to the formation of ionized layers by absorption of solar radiation in a simple atmosphere, assuming as an approximation that the earth is plane, the maximum rate of ion production in the layer, q_{max}, is proportional to $\cos\chi$, where χ is the sun's zenith distance. We may thus write

$$q_{max} = q_0 \cos\chi, \qquad \ldots\ldots(18)$$

which will be valid for the curved earth except for large values of χ. q_0 being the maximum rate of ion production when the sun is vertically overhead. Equation (17) now becomes

$$N_{e(max)} = \left\{\frac{q_0 \cos\chi}{\alpha'(1+\lambda)}\right\}^{1/2} \qquad \ldots\ldots(19)$$

The maximum electron concentration in a layer can be determined by measuring the frequency of radio waves which just penetrate the layer (Appleton, 1931). When this is done for the E layer it is found that within the limits of experimental error $N_{e\,(max)}$ is proportional to $(\cos\chi)^{1\cdot2}$ from soon after sunrise until near sunset. Also if the noon values of $N_{e\,(max)}$ for summer and winter are compared it is found that the $(\cos\chi)^{1\cdot2}$ law is obeyed, but in this case there is greater uncertainty on account of the variation of q_0 from year to year.

These experimental results teach us two important facts, viz.: (i) *The ionization density of the E region varies with the ionizing power of the sun's rays according to a recombination law*; and (ii). *during most of the day the rates of ionization and recombination are in equilibrium, there being no measurable asymmetry of the graph of ionization density relative to the time of maximum ion production.*

From (ii), as shown by Appleton (1937), we can set a lower limit to the recombination coefficient. The condition for equilibrium, from equation (14) above, is that $\frac{dN_e}{dt}$ should be small compared with $\frac{q}{1+\lambda}$ or $\alpha' N_e^2$; substituting observed values of $\frac{dN_e}{dt}$ and N_e^2, it is found that in order that $\frac{dN_e}{dt}$ should be less than, say, one-tenth of $\alpha' N_e^2$, α' must be greater than 10^{-8} cm^3 sec^{-1}.

If we wish to arrive at an actual figure for the E-region recombination coefficient we must examine the variation of N_e with time more closely near sunrise and sunset, where equilibrium does not prevail. For this we require to solve the differential equation

$$\frac{dN_e}{dt} = \frac{q(\chi, z)}{1+\lambda} - \alpha' N_e^2 \qquad \ldots\ldots(20)$$

for various values of α', and compare the solutions with the observed diurnal variation of N_e. A set of such solutions have been obtained by Wilkes (1939), using a differential analyser, and compared with observations of E-region $N_{e\,(max)}$ during undisturbed days. $q(\chi, z)$ is the Chapman function for the variation of q with χ and height z for a spherical earth, and was adjusted to fit the observations near mid-day. The value of α' which gave a good fit with observations made at Cambridge by Best, Farmer and Ratcliffe (1938) was about $1{\cdot}2 \times 10^{-8}$ cm^3 sec^{-1}, though a value slightly higher than this would give similar agreement.

Hulbert (1939) has made a similar comparison of theoretical curves for the diurnal variation of N_e with observations made at Washington, D.C., and Huancayo, Peru. The best fit was obtained with a value of α' of 2×10^{-8} cm^{3}sec^{-1}, which is in satisfactory agreement with the above figure.

The detailed study of N_e during an eclipse seems to offer the only possibility of the more accurate measurement of α for the E region. During eclipses the ionization density varies much more rapidly than in the normal diurnal cycle, and any lag of ionization density relative to the rate of ion production would therefore be more pronounced and would lead to a better determination of α. The records obtained at Washington, by Kirby, Gilliland, Berkner and Norton (1933, 34) duringthe eclipse of 31 August 1932, have therefore been examined by Wilkes (1939) and compared with theoretical N_e versus time curves for various

values of α. It was not possible to determine α with any accuracy from these records, but it can at least be stated that the eclipse results are not in disagreement with the above values of α.

During the night, if we assume that there is no source of ionization in the E region, the equation for the variation of N_e with time reduces to the simple form

$$\frac{dN_e}{dt} = -\alpha' N_e^2. \qquad \ldots\ldots(21)$$

or, integrating this, we have

$$N_e = \frac{1}{k+\alpha' t}, \qquad \ldots\ldots(22)$$

where t is time measured from sunset and k^{-1} is the value of N_e at $t=0$. If for α' we take the daytime value of about 2×10^{-8} cm^3 sec.$^{-1}$, this formula gives a figure for N_e in the early morning before sunrise which is smaller by a factor of 4 or 5 than the radio measurements indicate. We therefore conclude that either (i) there is some process of ionization, or rather of increase of electron concentration, which continues throughout the night, or (ii) the recombination coefficient is much smaller during the night, being, as Ratcliffe (1938) has estimated, about 4×10^{-9} cm^3 sec.$^{-1}$. There is considerable evidence that (i) is true during the first half of the night at least, for many observers have recorded a small but definite increase of E-region critical frequency after sunset, reaching a mild maximum towards midnight. Various suggestions, such as the action of supposed winds, atomic collision processes, and the contraction of the layer by cooling, have been advanced to explain this increase of ionization density. One possibility is that an ionization process, in addition to ultra-violet light from the sun, is continually active in the E region, but is normally swamped by the latter process during the day. This theory is supported by the fact that the occasional appearance of an "intense" or "abnormal" E ionization may be attributed to such an agency.

If any such process as the above continues throughout the night it is obvious that we cannot use the simple relation of equation (22) to determine the recombination coefficient; and there is then no evidence for a difference in the day and night values of α. The expression obtained in the last section, equation (16), for the effective recombination coefficient is

$$\alpha_0 + \lambda\alpha_1,$$

where

$$\lambda = \frac{\beta N_0}{\gamma N_0 + I}.$$

There is no reason to believe that any great change can occur in either α_0 or α_1; neither is it easy to see how λ could decrease at sunset as, if I, the rate of photo-ionization of negative ions, were appreciable compared with other detachment processes, λ should actually be greater at night.

(iii) *Comparison of E region recombination coefficient with laboratory measurements*

If we compare the effective recombination coefficient of 2×10^{-8} cm^3 sec.$^{-1}$ for day conditions with the expression $\alpha_0 + \lambda\alpha_1$, we observe that, since the

electron positive ion recombination coefficient cannot be much greater than 5×10^{-10} cm^3 sec.$^{-1}$, the α_e term may be neglected, and we are then left with the approximate relation

$$\lambda\alpha_i = 2 \times 10^{-8} \text{ cm}^3 \text{ sec.}^{-1}.$$

Now, in the pressure range from one atmosphere down to a few mm. Hg, the ion-ion recombination coefficient is pressure dependent, according to the Thomson theory, and takes place by a three-body process. It is therefore evident that, if the pressure is sufficiently reduced, recombination by way of two-body encounters, which is negligible at pressures of a few cm. Hg and greater, may be more effective than the three-body process. Such two-body processes, for obvious reasons, cannot be observed in laboratory experiments, but we may infer their existence since they are analogous to the process of electron transfer when a positive ion of a gas of high ionization potential like He collides with a normal atom of lower ionization potential. For example

$$\text{He}^+ + \text{A} \rightarrow \text{He} + \text{A}^+, \qquad \ldots\ldots(23)$$

a type of reaction which has been observed by Harnwell (1927) and by Smyth and Steuckelberg (1929). In the case of positive and negative ions colliding we have

$$\text{X}^+ + \text{Y}^- \rightarrow \text{X} + \text{Y}, \qquad \ldots\ldots(24)$$

and some of the recombination energy may be accounted for by excitation of the atoms X and Y after the collisions.

There is unfortunately, little information from which to calculate the probability of such a reaction, but when we remember that there is very good evidence that the recombination coefficient is, to the accuracy of observations, pressure independent in region *E*, we may conclude that the contribution of the two-body process is much greater than the contribution of the three-body process. If we adopt 10^{-3} mm. Hg for the pressure in the *E* region, we find that the α for three-body collisions is $8{\cdot}5 \times 10^{-12}$ cm^3 sec.$^{-1}$, and if we specify that the two-body contribution must be at least ten times this, we find that α_i is not less than 10^{-10} cm.^{3} sec.$^{-1}$, and may, in fact, be orders of magnitude greater than this. This lower limit is only dependable in so far as we are justified in assuming that the Thomson expression for the recombination coefficient, which it has only been possible to check from pressures of a few cm. Hg upwards, is reliable down to 10^{-3} mm. Hg pressure, and under conditions where the ions are atomic rather than molecular. It is, however, probable that such considerations will not account for a very great error, and the extrapolated value for α by three-body collisions should not be more than a factor of 2 in error.

(iv) *Recombination in the F region*

Our knowledge concerning the F_1 and F_2 layers is not nearly so detailed as for the *E* layer. F_2 differs from *E* in many respects. For example, the diurnal variation of ionization is much smaller than for the latter, and is subject to comparatively large fluctuations in intensity. Moreover, in contrast to conditions in region *E*, the ionization in F_2 lags behind the variations in ionizing agency, assuming the latter to be ultra-violet light from the sun. The presence of this

time lag, considering the high intensity of ionization, would indicate a small recombination coefficient in F_2. F_1, lying intermediate in height between E and F_2, also possesses intermediate characteristics, and might, during the day, be regarded as a simple region.

Observations on the critical frequencies for F_1 and F_2 during the partial solar eclipse at Washington in 1935 have been made by Kirkby, Gilliland and Judson (1936). On account of the ill-defined nature of the F_1 critical frequencies, the results for the F_1 layer were necessarily inconclusive, but there was some indication of a decrease in ionization density which seemed to lag behind the eclipse by about 20 minutes. For F_2 there was a better correlation between the decrease of ionization and the time of the eclipse, though there seemed to be a lag in this case also of about 9 or 10 minutes. By comparing curves for the fractional decrease of N_e and of N_e^2 with the fractional change of the exposed area of the sun's disc, they concluded that the curve for the change of N_e gave the best fit, and that the destructive process was therefore electron attachment. It must, however, be remembered that F_2 is usually subject to considerable fluctuations in ionic density, and since these observations were only for a partial eclipse in which the changes in the rate of ion production were not very great it might be difficult to distinguish between a recombination and an attachment law.

Assuming recombination to be the process by which the ionization in F_2 disappears, it is possible to estimate the recombination coefficient for a summer or equinox night in which, under quiet conditions, the electron concentration decreases in a regular fashion throughout most of the night. Under such conditions Appleton (1937) has determined the night-time recombination coefficient, and he had also devised another method of estimating the value of α during the day. Thus, in a series of experiments, a study of the F_2 critical frequency was made for nights during which the decrease in ionization was steady, with no signs of the nocturnal increase so often observed in winter. Typical results are given in the following table.

Nocturnal recombination coefficient in F_2 layer

Period	Recombination coefficient α 10^{-10} cm.^{3} sec.$^{-1}$
25/26 April 1933	3·7 to 6·0
11/12 April 1934	3·6
Average April 1934	2·5
16/17 May 1934	2·9
24/25 July 1934	3·4
Average September 1934	1·8
19/20 March 1935	2·8

Daytime values of the recombination coefficient were estimated from the asymmetry of the N_e versus time curves about mid-day. This was done by choosing two points from such a curve, one in the morning and one in the afternoon, such that the sun's zenith angle is the same in the two cases. Then, assuming that there is no appreciable alteration in the condition of the atmosphere,

the level of maximum ionization will be at the same height, and the rate of production of ions, q, per c.c. per sec. will be the same for the two cases. We may therefore write

$$\frac{dN_1}{dt} = q - \alpha N_1^2 \qquad \text{.....(25)}$$

and

$$\frac{dN_2}{dt} = q - \alpha N_2^2.$$

Since N_1, N_2 and their time derivations are measurable quantities, α may be calculated from these equations. By this method Appleton finds that for a midsummer day and an equinox day the respective values of α are $8{\cdot}7 \times 10^{-11}$ and $8{\cdot}1 \times 10^{-11}$ cm^3 sec^{-1}.

Taking 10^{-10} cm^3 sec.$^{-1}$ as a round figure for the recombination coefficient in F_2, we may write

$$\alpha_e + \lambda\alpha_i = 10^{-10}\,\text{cm}^3\,\text{sec}^{-1}. \qquad \text{.....(26)}$$

It is of interest to note that attempted laboratory measurements of electron and positive-ion recombination coefficients at low pressure have indicated a value of approximately 10^{-10} cm^3 sec.$^{-1}$ for α_e. If this is significant, it means that the concentration of negative ions in F_2 must be very small. But it is important to note that it is far from clear what mechanism is responsible for the above value observed in laboratory experiments. Obviously much more information is required, both by radio exploration and laboratory experiments, before these questions can be settled.

REFERENCES

Appleton, E. V., 1931. *Nature, Lond.*, **127**, 197.
Appleton, E. V., 1933. *Proc. Roy. Soc.* A, **141**, 697.
Appleton, E. V., 1937. *Proc. Roy. Soc.* A, **162**, 451.
Appleton, E. V. and Naismith, R., 1932. *Proc. Roy. Soc.* A, **137**, 36.
Appleton, E. V. and Naismith, R., 1935. *Proc. Roy. Soc.* A, **150**, 685.
Appleton, E. V. and Sayers, J., 1938. *Sixth Gen. Assembly U.R.S.I. Proc.* p. 272.
Bates, D. R., Buckingham, R. A., Massey, H. S. W. and Unwin, J. J., 1939. *Proc. Roy. Soc.* A, **170**, 322.
Best, J. E., Farmer, F. T. and Ratcliffe, J. A., 1938. *Proc. Roy. Soc.* A, **164**, 96.
Bradbury, N. E., 1933. *Phys. Rev.* **44**, 883.
Chapman, S., 1931. *Proc. Roy. Soc.* A, **132**, 353.
Cravath, A. M., 1929. *Phys. Rev.* **33**, 605.
Gish, O. H. and Sherman, K. L., 1936. *Bull. Int. Ass. Terr. Mag. Elect.* no. 10, p. 382.
Harnwell, G. P., 1927. *Phys. Rev.* **29**, 683 and 830.
Hulbert, E. O., 1939 a. *Phys. Rev.* **55**, 639.
Hulbert, E. O., 1939 b. *Phys. Rev.* **55**, 646.
Kirby, S. S., Berkner, L. V., Gilliland, T. R. and Norton, K. A., 1933 and 1934. *Bur. Stand. J. Res., Wash.*, **11**, 829.
Kirby, S. S., Gilliland, T. R. and Judson, E. B., 1936. *Proc. Inst. Radio Engrs., N.Y.*, **24**, 1027.
Martyn, D. F. and Pulley, O. O., 1936. *Proc. Roy. Soc.* A, **154**, 455.
Massey, H. S. W., 1937. *Proc. Roy. Soc.* A, **163**, 542.
Massey, H. S. W., 1938. *Negative Ions* (Cambridge : University Press).
Sayers, J., 1938. *Proc. Roy. Soc.* A, **169**, 83.
Smyth, H. D. and Steuckelberg, E. C. G., 1929. *Phys. Rev.* **32**, 779.
Stevens, A. W. and Anderson, O. A., 1938. *Nature, Lond.*, **141**, 270.
Tukada, T., 1937. *Rep. Rad. Res., Japan*, **7**, 121.
Wilkes, M. V., 1939. *Proc. Phys. Soc.* **51**, 138.

THE PROPERTIES OF NEUTRAL AND IONIZED ATOMIC OXYGEN AND THEIR INFLUENCE ON THE UPPER ATMOSPHERE

By H. S. W. MASSEY, F.R.S. and D. R. BATES,
University College, London

§ 1. INTRODUCTION

It has been shown by Chapman (1930a, 30b) that the photo-dissociation of molecular oxygen by solar radiation has a very important consequence. At great heights in the atmosphere all the oxygen is present in the atomic form and some intermediate transition region can be specified such that, at lower levels, the oxygen is predominantly molecular, at higher levels, predominantly atomic. Detailed investigation (Wulf and Deming, 1938) shows that this transition region probably occurs at a height of about 100 kilometres. As this is near the level of the E layer and is below that of the F_1 and F_2 layers, it is clear that a knowledge of the properties of atomic oxygen, both neutral and ionized, is essential for the understanding of ionospheric phenomena. As the possibilities of experimental investigation of most of these properties are very limited, it is necessary to rely very largely on theory. It is the primary purpose of this *Report* to summarize the information at present available, mainly from theoretical sources, concerning the properties of O, O^+ and O^-. In the concluding section, however, a brief account of applications to the upper atmosphere is given.

The first section deals with the electronic structure, energy levels and line spectra of O, O^+ and O^-, considering, in particular, the metastable levels and the so-called forbidden lines. In the second section collision processes involving the neutral atom—photoionization (continuous absorption), elastic scattering of electrons by O atoms, and deactivation of metastable O atoms by electron collision are discussed. The recombination of electrons with O^+ ions is the subject of the third section, while the fourth is concerned with modes of formation and destruction of O^- ions. Mutual neutralization on collision of O^+ and O^- is considered in the fifth section. The concluding section deals briefly with applications to the theory of the origin of the ionized layers, of the rate of ionospheric recombination and of the origin of night-sky light.

§ 2. STRUCTURE, ENERGY LEVELS AND LINE SPECTRA OF O, O^+ AND O^- *

2.11 *Energy levels of O—the metastable terms*

Full details of the energy levels of this atom, as determined from analysis of its spectrum, have been given in *Atomic Energy States*, by Bacher and Goudsmit,

* An excellent detailed account of the spectra of O and O^+ has been given by Nicolet (1939).

Table 1. The deeper energy levels of O, O^+ and O^-

	Configuration	Term	Remarks
Neutral atom, O	$(1s)^2\,(2s)^2\,(2p)^4$	^3P	Ground term. Ionization potential = 13·55 v.
		^1D	Metastable term, 1·96 ev. above ^3P
		^1S	Metastable term, 4·17 ev. above ^3P
	$(1s)^2\,(2s)^2\,(2p)^3\,3s$	$^5S^0$	9·10 ev. above ^3P
Positive ion, O^+	$(1s)^2\,(2s)^2\,(2p)^3$	$^4S^0$	Ground term. Ionization potential = 34·93 v.
		$^2D^0$	Metastable term, 3·31 ev. above $^4S^0$
		$^2P^0$	Metastable term, 4·99 ev. above $^4S^0$
	$(1s)^2\,(2s)\,(2p)^4$	^4P	14·78 ev. above $^4S^0$
Negative ion, O^-	$(1s)^2\,(2s)^2\,(2p)^5$	$^2P^0$	Ground term. Detachment potential = 2·2 v.
	$(1s)^2\,(2s)^2\,(2p)^4\,3s$	^4P	The energies of these terms are not known with certainty, but they may be nearly equal to that of the ground level of O
		^2P	
	$(1s)^2\,2s\,(2p)^4$	^2S	

so it is only necessary to summarize a few essential facts concerning them. These are given in table 1. It will be noted that two metastable terms arise from the lowest configuration.

2.12 *The "auroral", "trans-auroral" and "nebular" O lines**

In the upper air the rate of deactivation of an excited atom by collision is so low that even a metastable state of long radiative lifetime has a considerable chance of emitting a spectrum line by a "forbidden" transition. The three transitions concerned in atomic oxygen are those between ^1S, ^1D and ^3P terms.

Transitions between ^1S and ^1D give rise to the famous green line, a single line of wavelength 5577 A. The finite probability of the transition arises from the effective electric quadrupole moment associated with it, even when no deviation from Russell-Saunders coupling is assumed. This has been established by observation † of the Zeeman effect of the line produced experimentally Theoretical values for the lifetime of the ^1S term towards this transition were first given by Bartlett (1929) using a rough approximation. A more rigorous treatment was then given by Stevenson (1932) and a final detailed analysis by Condon (1934). The latter author gives the transition probability as 2·0/sec. for emission of the green line.

Transitions between ^1D and ^3P give the nebular red doublet of wavelengths 6300 and 6364 A. respectively. The associated transition probabilities are much smaller, being 7·5 and $2{\cdot}5 \times 10^{-3}$/sec. respectively, as the lines only arise owing to departures from Russell-Saunders coupling, i.e. by spin-orbital interaction.

Transitions between ^1S and ^3P give the trans-auroral lines. Both of these can only arise from spin-orbital interaction, but one is due to magnetic-dipole,

* The names "auroral", "trans-auroral" and "nebular" were introduced by Boyce, Menzel and Payne (1933).

† Frerichs and Campbell (1930).

the other to electric-quadrupole, emission. As a consequence, the former, of wavelength 2972 A., is much stronger, the transition probability being 0·18/sec. as against $1{\cdot}6 \times 10^{-4}$ sec. for the quadrupole line (2958 A.).

2.21 *Energy levels and metastable terms of O⁺*

Table 1 summarizes the properties of the relevant low-lying terms of O^+. As for the neutral atom, the two excited terms associated with the ground configuration give rise to metastable states.

2.22 *Forbidden lines of O⁺*

The theory of these lines has also been discussed by Condon (1934). His calculated transition probabilities are given in table 2, which also summarizes the various forbidden lines which can occur, with their wavelengths.

Table 2. The forbidden spectral lines of oxygen

	Transition		Wavelength (A.)	Probability (sec.$^{-1}$)
Neutral atom, O	$^3P_0 - {}^1D_2$	nebular red	6392	0
	$^3P_1 - {}^1D_2$		6364	$2{\cdot}5(10)^{-3}$
	$^3P_2 - {}^1D_2$		6300	$7{\cdot}5(10)^{-3}$
	$^1D_2 - {}^1S_0$	auroral green	5577	2·0
	$^3P_0 - {}^1S_0$	trans-auroral	2978	0
	$^3P_1 - {}^1S_0$		2972	$1{\cdot}8(10)^{-1}$
	$^3P_2 - {}^1S_0$		2958	$1{\cdot}6(10)^{-4}$
Positive ion, O^+	$^2D_{3/2} - {}^2P_{3/2}$	auroral multiplet	7331	1·7
	$^2D_{3/2} - {}^2P_{1/2}$		7330	3·3
	$^2D_{5/2} - {}^2P_{3/2}$		7320	2·3
	$^2D_{5/2} - {}^2P_{1/2}$		7319	1·0
	$^4S_{3/2} - {}^2D_{5/2}$	nebular pair	3729	$3{\cdot}7(10)^{-5}$
	$^4S_{3/2} - {}^2D_{3/2}$		3726	$2{\cdot}4(10)^{-5}$
	$^4S_{3/2} - {}^2P_{3/2}$	trans-auroral	2470	$4{\cdot}1(10)^{-2}$
	$^4S_{3/2} - {}^2P_{1/2}$		2470	$4{\cdot}9(10)^{-2}$
Negative ion, O^-	There may be a discrete spectral line in the region 5000–6000 A. arising, with low probability, from $(2p)^4\ 3s\ {}^4P$ or with high probability (an allowed transition) from $(2p)^4\ 3s\ {}^2P$, $2s(2p)^6\ {}^2S$. (See table 1.)			

2.3 *Negative ions of atomic oxygen*

The existence of stable O^- ions is established beyond doubt. They have frequently been observed in mass spectrographic analysis of negative rays (see, for example, Tüxen, 1936); and Lozier (1934), by studying the energetics of such reactions as

$$O_2 + e \rightarrow O + O^-, \qquad \ldots\ldots(1)$$

using a special design of spectrograph, has determined the electron affinity as $2{\cdot}2 \pm 0{\cdot}2$ ev. This value has been confirmed by the later work of Hagstrum and Tate (1941). Furthermore, Hartree, Hartree and Swirles (1939) have shown theoretically that O^- must be stable. They obtain a value of 0·50 ev. for the electron affinity by a method which, by its nature, must lead to an under-estimate.

The configuration of the ground state of O^- must be $(1s)^2(2s)^2(2p)^5$, which can give rise to only one, a $^2P^0$, term. The number of discrete excited states of a negative ion is known from theoretical arguments to be limited (Massey, 1938), and it is probable that in most cases only the ground level is stable. Evidence from the experiments of Lozier (1934) and Hagstrum and Tate (1941) in CO and NO has indicated the possible existence of a stable excited state of O^- with very small binding energy. If such a state exists, it must belong to either the $(1s)^2(2s)^2(2p)^4 3s$ or $(1s)^2(2s)(2p)^6$ configuration, and there are empirical arguments in favour of the former of these. Bates and Massey (1943) have discussed theoretically the conditions under which this configuration could give a stable excited term. They introduce an effective polarizability p of the core by the $3s$ electron as an adjustable parameter and find the values p_1, p_2 of p, necessary just to introduce stable ^4P and ^2P terms respectively. These values turn out to be over twice as large as that deduced from theory and from the refractivity of O_2. As the interpretation of the experiments is not entirely unambiguous, the existence of the stable excited level is open to doubt. However, if it were to exist with very low binding energy and belong to the configuration considered, the probabilities of radiative attachment and detachment of electrons respectively to and from O atoms would be drastically affected as well as that of their elastic scattering.* In all these cases, calculations have, therefore, been carried out for various assumed values of the parameter p, including those, p_1, p_2, introducing stable excited levels.

§ 3. COLLISION PROCESSES ASSOCIATED WITH NEUTRAL O ATOMS

3.1 *Photoionization*

The ionization of O atoms by radiation plays a very important part in shaping upper atmospheric phenomena. Removal of a $2p$ electron from the ground state of O leaves an O^+ ion in either a $^4S^0$, $^2D^0$ or $^2P^0$ state, the energy required being 13·55, 16·86 and 18·54 ev. respectively. To remove an s electron requires a much greater energy (28·33 ev. or more), so the continuous absorption of radiation of quantum energy less than 25 ev. involves ejection of a $2p$ electron only. Let $k(^4S^0)$, $k(^2D^0)$, $k(^2P^0)$ be partial cross-sections for absorption, leaving O^+ ions in $^4S^0$, $^2D^0$, $^2P^0$ states respectively, so the total absorption cross-section is given by

$$k = k(^4S^0) + k(^2D^0) + k(^2P^0). \qquad \ldots\ldots(2)$$

Quantum theory has been applied by Saha and Rai (1938), by Bates, Buckingham, Massey and Unwin (1939) and by Yamanouchi and Kotani (1940) to calculate k. Their results show that, at quantum energies above the appropriate ionization energy and less than 25 ev., $k(^4S)$, $k(^2D)$, $k(^2P)$ are approximately independent of frequency and we have

$$\begin{aligned} k &= 4{\cdot}5 \times 10^{-18}\,\text{cm}^2, && 13{\cdot}55\,\text{ev.} \leqslant h\nu \leqslant 16{\cdot}86\,\text{ev.} \\ &= 11 \times 10^{-18}\,\text{cm}^2, && 16{\cdot}86\,\text{ev.} \leqslant h\nu \leqslant 18{\cdot}54\,\text{ev.} \\ &= 16 \times 10^{-18}\,\text{cm}^2, && 18{\cdot}54\,\text{ev.} \leqslant h\nu \leqslant \textit{circa}\ 25\,\text{ev.} \end{aligned}$$

* Resonance effects appear in the elastic scattering when p is such as just to introduce either a stable ^4P or ^2P term; but, in attachment and detachment, only the latter (slightly larger) value produces resonance.

It will be noted that the absolute value of k is not abnormally small, and in such circumstances the quantum calculations can be expected to give accurate results (Bates and Massey, 1941). The other important feature is that k does not fall off rapidly with increasing quantum energy, but remains high over the entire energy range concerned. The importance of this will be pointed out in §6.

3.2 *Elastic scattering of slow electrons by atomic oxygen*

The effective target area presented by oxygen atoms to electrons with energies corresponding to temperatures up to 1000° K. depends very much on the atomic field force at large nuclear distances. In particular the attraction due to the polarization of the core is very important. It has been noted in §2.3 that this force is also the most important in determining the stability or otherwise of any excited states of the O^- ion, and in table 4 the values p_1, p_2 of the effective polarizability, p, of the atomic core sufficient just to introduce stable $(1s)^2(2s)^2(2p)^4\,3s\,{}^4P$ and ${}^2P$ terms, respectively, are given. In default of any more definite method, and until the existence of a stable excited state is confirmed, or otherwise, the elastic scattering must be calculated for a variety of values of p ranging from zero to just greater than p_2. If the actual long-range force is best represented by a value of p near p_1 or p_2, the cross-section for elastic scattering of slow electrons must be very large owing to resonance effects. Thus the mean collision coefficient * for electrons with Maxwellian distribution of velocities at normal temperatures may be as high as 10^{-4} cm^3/sec. Calculations covering the whole range from $p=0$ to $p=p_2$ have not been completed, as at some intermediate values of p resonance effects due to incident p electrons probably occur and have not yet been taken into account. It is clear that more information concerning the true value of p is required before definite collision frequencies for impact of slow electrons with atomic oxygen can be derived. In the meantime, the possibility of the large values arising if $p \sim p_1$ or p_2 cannot be overlooked.

3.3 *Deactivation of metastable oxygen atoms by impact*

Metastable atoms may transfer their excitation energy either by impact with a neutral molecule or an electron. If we restrict ourselves to collisions in which the particles concerned have thermal energies only, corresponding to temperatures of 1000° K. or less, deactivation must occur by second-kind transfer, i.e. the metastable atoms must return to the ground state and the colliding body gain the surplus energy either in potential or kinetic form. Quite large effective cross-sections can, therefore, be expected. Yamanouchi, Inui and Amemiya (1940) have calculated cross-sections for this type of deactivation by electrons, but the numerical accuracy of the theory is doubtful owing to the sensitiveness of the answer to changes in the atomic magnitudes concerned.

§4. RECOMBINATION OF ELECTRONS WITH O^+ IONS

In order that an electron should recombine with an O^+ ion to reform an O atom in the normal or an excited state, the surplus energy of the electron must

* The collision frequency per electron is obtained from the collision coefficient by multiplication by the number of atoms/c.c.

be got rid of in some way. If the pressure is so low that no third body is likely to be sufficiently close to receive the energy, only two modes of energy disposal are possible. Either it is directly radiated or else is communicated to a second electron in the O^+ ion, raising it to an excited orbital. In the latter case, the state in which the incident and second electron are both in excited orbitals is not stable and can either revert to the initial condition by re-emission of an electron or, by emission of radiation, undergo a transition to a stable state of the neutral atom, leading thus to recombination. This process is the inverse of that known as auto-ionization, and its contribution cannot be ignored. Since it depends on interaction between two electrons we refer to it hence-forward as dielectronic recombination. We now consider separately the contributions from the two modes of energy disposal.

4.1 *Radiative recombination*

The contribution to the electron-recombination coefficient due to this process has been calculated in detail by Bates, Buckingham, Massey and Unwin (1939).* Their results are given in table 3 for a number of electron energies, expressed in terms of a temperature in the usual way. Contributions arising from re-combination to the ground state and to excited states are given separately. It

Table 3. Radiative recombination coefficients for atomic oxygen

α in 10^{-12} cm^3/sec.

Temperature (°K.)	Contribution from ground state	Contribution from excited states	Total
250	$0{\cdot}78_2$	2·91	3·69
500	$0{\cdot}55_4$	1·84	2·39
1000	$0{\cdot}39_9$	1·09	1·48
2000	$0{\cdot}27_8$	$0{\cdot}61_4$	$0{\cdot}89_2$
4000	$0{\cdot}19_8$	$0{\cdot}33_7$	$0{\cdot}53_5$
8000	$0{\cdot}13_8$	$0{\cdot}18_9$	$0{\cdot}32_7$

will be seen that, even at the lowest electron energies given, corresponding to a temperature of 250° K., the total recombination coefficient from radiative re-combination is only $3{\cdot}7 \times 10^{-12}$ cm^3/sec.

4.2 *Inverse auto-ionization or dielectronic recombination*

If E_s is the energy excess of a term of a configuration of O, involving two excited electrons, above the energy of the ground level of O^+, and τ_s the lifetime of the term towards a radiative transition to a stable level, the recombination coefficient due to this process, summed over all possible doubly excited levels, is

$$h^3(2\pi mkT)^{-3/2}\Sigma_s(w_1 2w_2)\,\tau_s^{-1}e^{-E_s/kT}, \qquad \ldots\ldots(3)$$

* The values given in the third column of table 2 of that paper should be multiplied by 2, owing to the omission of a weighting factor (Bates and Massey, 1943). The fifth column must then be appropriately modified to give the values shown in table 3 of the present paper.

Table 4. Rates of radiative attachment and photo-detachment

Value of polarizability parameter p (see § 2.3) in atomic units	Rate of radiative attachment per electron per atom per cm³ in 10^{-13} cm³/sec. Electron temperature				Rate of photo-detachment per negative ion due to solar radiation at temperature 6000° K. in sec.$^{-1}$.	Ratio of attachment and detachment rates in 10^{-13} cm³ Electron temperature			
	250° K.	500° K.	1000° K.	2000° K.		250° K.	500° K.	1000° K.	2000° K.
0	$7{\cdot}5_0 \times 10^{-3}$	$7{\cdot}7_1 \times 10^{-3}$	$8{\cdot}0_3 \times 10^{-3}$	$8{\cdot}3_3 \times 10^{-3}$	$0{\cdot}27_4$	$2{\cdot}7_1 \times 10^{-2}$	$2{\cdot}8_0 \times 10^{-2}$	$2{\cdot}9_0 \times 10^{-2}$	$3{\cdot}0_0 \times 10^{-2}$
5·7 (predicted polarizability)	$1{\cdot}1_1 \times 10^{-2}$	$1{\cdot}1_2 \times 10^{-2}$	$1{\cdot}1_4 \times 10^{-2}$	$1{\cdot}1_5 \times 10^{-2}$	$0{\cdot}34_8$	$3{\cdot}2_1 \times 10^{-2}$	$3{\cdot}3_0 \times 10^{-2}$	$3{\cdot}3_0 \times 10^{-2}$	$3{\cdot}3_0 \times 10^{-2}$
14·0 (required just to introduce stable ^4P term)	$3{\cdot}9_8 \times 10^{-2}$	$3{\cdot}5_4 \times 10^{-2}$	$3{\cdot}1_6 \times 10^{-2}$	$2{\cdot}6_3 \times 10^{-2}$	$0{\cdot}54_7$	$7{\cdot}3_1 \times 10^{-2}$	$6{\cdot}5_1 \times 10^{-2}$	$5{\cdot}8_1 \times 10^{-2}$	$4{\cdot}8_1 \times 10^{-2}$
17·6 (required just to introduce stable ^2P term)	$5{\cdot}8_5 \times 10^{-1}$	$2{\cdot}9_5 \times 10^{-1}$	$1{\cdot}5_3 \times 10^{-1}$	$7{\cdot}9_9 \times 10^{-2}$	$0{\cdot}85_4$	$6{\cdot}8_1 \times 10^{-1}$	$3{\cdot}4_1 \times 10^{-1}$	$1{\cdot}8_1 \times 10^{-1}$	$9{\cdot}2_1 \times 10^{-2}$

In calculating rate of photo-detachment the dilution factor for the solar radiation is taken as $5{\cdot}43_5 \times 10^{-6}$

provided the incident electrons have a Maxwellian energy distribution characteristic of the temperature T. w_1 and w_2 are the respective weights of the doubly excited states and of the ground states of O^+.

In atomic oxygen a study of the energy levels shows that only terms based on the lower states of O^+ can contribute appreciably, and (3) then shows that it is unlikely that the recombination coefficient arising from this source can be greater than 10^{-12} cm^3/sec.

It will be seen that the coefficient for direct recombination of electrons with positive ions would not be expected to exceed 5×10^{-13} cm.^{3}/sec. Discharge-tube experiments carried out by Kenty (1928) in argon, and by Mohler (1937) in caesium and mercury, have indicated values as high as 2×10^{-10} cm^3/sec. For these heavier ions the contribution from dielectronic recombination may be greater than for O^+, but it is unlikely that this would be sufficient to explain the discrepancy. Further investigation is clearly required.

§ 5. PRODUCTION AND DESTRUCTION OF O^- IONS

In discussing processes by which O^- ions are produced or destroyed it is convenient to consider separately each attachment process, together with the corresponding inverse process which leads to detachment, the probabilities of the two being related by the principle of detailed balancing.

As for recombination, attachment of an electron to form O^- can only occur if the surplus energy is disposed of. This can be done either by radiation, by communication to a second atomic electron involving inverse "auto-detachment" or dielectronic attachment (analogous to inverse auto-ionization or dielectronic recombination), or, if the electron impinges on a molecule containing O atoms, the energy can be absorbed in breaking the molecular bond and producing relative motion between the resulting O^- ion and the residual molecule (dissociative attachment).

The probabilities of the different processes and of their inverse reactions have been discussed by Bates and Massey (1943) and the results they obtain are as follows.

5.1 *Radiative attachment and photo-detachment*

As noted in § 2.3, the rates of these processes cannot be calculated absolutely, for, if the effective polarizability of the atomic core by a slow electron is in the neighbourhood of p_2 (see § 2.3), resonance effects influence radiative attachment and photo-detachment as well as elastic scattering. In table 4 mean attachment and detachment rates are given, calculated for values of p between $p = 0$ and a value slightly greater than p_2. The mean for the attachment rates is taken for a swarm of electrons with Maxwellian distribution of velocities at the temperature specified, that for detachment for quanta having a frequency distribution characteristic of a black body at temperature 6000° K.* The ratio of the two rates under the different conditions and for the various assumed values of p is also given in table 4, as it plays an important rôle in the theory of ionic equilibrium in the ionosphere.

* The intensity is "diluted" by a factor $5{\cdot}43_8 \times 10^{-6}$, the ratio to 4π of the solid angle subtended at the Earth by the Sun.

Table 5. Rates of dielectronic attachment and auto-detachment

Energy excess of doubly excited state in ev.	Rate of dielectronic attachment per electron per atom per $cm.^3$ in 10^{-13} $cm.^3$/sec. Electron temperatures				Rate of auto-detachment per negative ion due to solar radiation at temperature 6000° K. in $sec.^{-1}$	Ratio of attachment and detachment rates in 10^{-13} $cm.^3$ Electron temperatures			
	250° K.	500° K.	1000° K.	2000° K.		250° K.	500° K.	1000° K.	2000° K.
0	$1{\cdot}1_4$	$4{\cdot}0_6\times10^{-1}$	$1{\cdot}4_3\times10^{-1}$	$5{\cdot}0_6\times10^{-2}$	$2{\cdot}5\times10^{-1}$	$4{\cdot}4_0$	$1{\cdot}6_0$	$5{\cdot}6\times10^{-1}$	$1{\cdot}9_9\times10^{-1}$
0·25	$1{\cdot}0\times10^{-6}$	$1{\cdot}2\times10^{-3}$	$7{\cdot}9\times10^{-3}$	$1{\cdot}1_9\times10^{-2}$	$1{\cdot}5_6\times10^{-1}$	7×10^{-5}	8×10^{-3}	$5{\cdot}0\times10^{-2}$	$7{\cdot}6\times10^{-2}$
0·50	$1{\cdot}0\times10^{-10}$	$3{\cdot}7\times10^{-6}$	$4{\cdot}3\times10^{-4}$	$2{\cdot}8\times10^{-3}$	$9{\cdot}6\times10^{-2}$	1×10^{-6}	4×10^{-5}	$4{\cdot}0\times10^{-3}$	$2{\cdot}9\times10^{-2}$
1·00	$8{\cdot}0\times10^{-21}$	$3{\cdot}0\times10^{-11}$	$1{\cdot}3\times10^{-6}$	$1{\cdot}5\times10^{-4}$	$3{\cdot}6\times10^{-3}$	2×10^{-10}	9×10^{-10}	$4{\cdot}0\times10^{-5}$	$4{\cdot}0\times10^{-3}$

In calculating rate of auto-detachment the dilution factor for the solar radiation is taken as $5{\cdot}43_8\times10^{-6}$.

5.2 *Dielectronic attachment and its inverse*

A formula of the form (3) gives the rate of this type of process to which only one doubly excited configuration can possibly contribute. This is the configuration $(1s)^2 2s(2p)^4$, and, even for it to be effective, it must include a term with energy E_0+E_1, slightly in excess of that, E_0, of the ground state of O. In this case the first stage involves capture of the incident electron releasing energy which excites an atomic electron from a $2s$ to a $2p$ orbital. Subsequently one of the $2p$ electrons returns to a $2s$ orbital by a radiative switch. In table 5 mean attachment and detachment rates arising from this process are given for various assumed values of the energy E_1. The absolute value of the attachment rate depends rather markedly on the velocity distribution of the electrons, and this must be remembered in applying the results. A Maxwellian distribution was assumed in deriving the values given in table 5. The radiative lifetime of the excited ion first formed on impact was taken to be 10^{-7} sec. For other lifetimes, τ, of the same order the attachment rate can be taken as inversely proportional to τ.

There is, unfortunately, as yet no evidence available as to the true magnitude of E_1.

5.3 *Dissociative attachment and its inverse*

O^- ions are produced by electron impact with O_2 molecules by such reactions as

$$O_2+e \rightarrow O+O^-, \qquad \ldots\ldots(4)$$

but the electron energies involved are considerably in excess of those likely to prevail in the upper air. The lowest electron energy at which (4) has been observed (Lozier, 1934) is 2·9 ev. It is unlikely, therefore, that dissociative attachment contributes appreciably to negative ion formation in the upper air,* but the inverse process, associative detachment, is likely to be important as providing a source of free electrons even in the absence of radiation, i.e. at night (Martyn and Pulley, 1936). The following reactions are energetically possible:

$$O^-+O \rightarrow O_2+e, \qquad \ldots\ldots(5)$$

$$O^-+N_2 \rightarrow N_2O+e, \qquad \ldots\ldots(6)$$

but

$$O^-+O_2 \rightarrow O_3+e \qquad \ldots\ldots(7)$$

is not, in the upper air.

In estimating the rates of the reactions (5) and (6) we must remember that (5) is not strictly the inverse of (4), as in (4) the O_2 molecules are in their ground vibrational state, whereas in (5) they may be formed with a high degree of vibrational excitation. Thus the application of the principle of detailed balancing to derive the rates of the detachment reaction (5) from the observed cross-section for the reaction (4) gives only a lower limit to the rate. From the experiments of Lozier (1934) and of Hagstrum and Tate (1941) this lower limit may be derived as between 10^{-21} and 10^{-23} cm^2. It is not possible to say how much this should be increased to allow for formation of vibrationally excited O_2, but a factor of at least 10 probably arises in this way, and larger factors cannot be excluded.

* It is just possible that the inverse of (7) below, which can occur with very slow electrons, may be of some importance, but the proportion of ozone present in the ionosphere is probably very small.

§6. MUTUAL NEUTRALIZATION OF O^+ AND O^- IONS ON IMPACT

When positive and negative ions collide there is a finite chance that an electron leaks across from the negative ion and occupies a vacant level in the positive ion. The result is the production of two O atoms, both of which may be in excited states, viz.,

$$O^+ + O^- \rightarrow O' + O''. \qquad \ldots\ldots(8)$$

Owing to their charges the ions may react effectively at very great distances, provided the "transition" interaction energy concerned, the magnitude of which determines the chance of reaction, is large enough. Furthermore, if the relative velocity of the ions at infinite separation is very small, corresponding, say, to ions at ordinary temperature, the reaction is essentially a collision of the second kind, and its rate is weighted up as a consequence. The various possible reactions of the type (8) which can arise have been discussed by Bates and Massey (1943). They were able to show that it is quite likely that, for one of these, the effective coefficient will be as large as 10^{-7}–10^{-8} cm^3/sec. The theory is not sufficiently definite to *prove* that such a high value will occur for any particular case, but it shows clearly that there are conditions which are probably met in some reaction which could give rise to it. On the other hand, a coefficient as large as 10^{-6} cm^3/sec. is most unlikely. For most reactions much lower coefficients are to be expected.

§7. APPLICATIONS TO THE IONOSPHERE

7.1 *Layer formation*

As it seems probable that the rate of loss of electrons follows the square recombination law (Appleton and Naismith, 1935), the layers of maximum ion density will be those of maximum ion production during daylight. Chapman (1931) has given a theory of the variation of ion production with height which shows that, for a particular absorbing constituent, the maximum rate of production occurs at a height where the number of molecules/cm^3 is $(1/kH)\cos\chi$, k being the absorption coefficient, H the partial scale height for the constituent concerned and χ the solar zenith distance. Considering ions produced from atomic oxygen by absorption of solar radiation of quantum energy between 13·5 and 15 ev., the value of k given in §3.1, with a scale height in the neighbourhood of 40 km., leads to an atomic oxygen concentration of less than 10^{11}/cm^3 at the level of maximum ion production. This concentration is so low that the level must be well above the E region, and can probably be identified with F_1. If this is so, all radiation with quantum energy greater than 13.5 ev., up to 25 ev. at least, must be strongly absorbed by atomic oxygen before region E is reached. This follows because the absorption coefficient does not fall as the quantum energy increases up to 25 ev. (see § 3.1). Hence it would appear that the E region can only be due to ionization of molecular oxygen, this being the only atmospheric constituent which can be ionized by radiation with quantum energy less than 13·5 ev.

Developing this in detail leads to a consistent theory of the origin of the layers.

7.2 *Ionospheric recombination*

The observed recombination coefficients in the E and F layers are many times greater than can be ascribed to direct radiative recombination of electrons and ions. Referring to table 3, the contribution to the recombination coefficient from this source is not likely to be greater than 5×10^{-12} cm^3/sec., whereas the mean observed values are 2×10^{-8} cm^3/sec. for E, between 10^{-8} and 10^{-9} cm^3/sec. for F_1 (daytime only) and 10^{-10} cm^3/sec. for F_2. To explain this discrepancy the influence of negative ions must be considered. It has been shown (Appleton and Sayers, 1938) that, if the ratio of negative ion to electron density at any time is λ, then the effective recombination coefficient for electrons can be written

$$\alpha = \alpha_e + \lambda\alpha_i, \qquad \ldots\ldots(9)$$

where α_e is the coefficient for direct electron-positive ion recombination and α_i that for negative ion-positive ion recombination or mutual neutralization. In the daytime, the maximum possible value λ_m of λ may be derived from the fourth columns of tables 4 and 5, since then we have to a close approximation

$$\lambda_m = \frac{\text{Rate of attachment/electron}}{\text{Rate of photo-detachment/negative ion}}. \qquad \ldots\ldots(10)$$

Substitution of numerical values shows that λ_m is probably between 1/20 and 1/2 for region E, assuming 10^{13} atoms/cm. at this level, and about 5 times smaller for region F_1. Turning now to § 6, it was noted there that the coefficient for mutual neutralization of O^+ and O^- may be as large as 10^{-7} cm^3/sec. Taking this in conjunction with the values of λ_m given above, substituting in (9) gives values of α of the correct order in daytime. At night in region E the observed value of α does not increase greatly, showing that the detachment rate due to reactions of the type (5) and (6) must be comparable with photo-detachment which prevails during the day. Examination of the numerical values concerned shows that the cross-section required for these reactions is then about 10^{-18} cm^2 under E-region conditions, and this, though high, is not impossible (see § 5.3). However, much remains to be cleared up before a completely satisfactory account of ionospheric recombination can be given. We have only indicated here how the information summarized in this *Report* may be utilized in this connection.

7.3 *Origin of night-sky radiation* *

The forbidden auroral green line of atomic oxygen is the source of the strongest night-sky radiation. The intensity has been estimated by Rayleigh (1930) as corresponding to 10^8 transitions per sec. in a column of the atmosphere of 1 sq. cm. base. Using the transition probability given in table 2, this implies that a concentration of metastable atoms of order 10–100 per cm^3 is maintained throughout the night. The source of these atoms presents a difficult problem which must be studied in relation to the observed variation of the green-line intensity with height, latitude and season and to the intensity of other nocturnal radiation. No definite solution can be advanced here, but the mutual neutralization process (8) must be considered as one possibility, associated with the

* For a detailed account of this radiation see Déjardin (1936), Chapman (1937), Elvey, Swings and Linke (1940), and the *Report* in this volume by Pearse.

ionized layers, which may provide an adequate supply of the metastable atoms concerned.

REFERENCES

APPLETON, E. V. and NAISMITH, R., 1935. *Proc. Roy. Soc.* A, **150**, 685.

APPLETON, E. V. and SAYERS, J., 1938. *Union Radio Scient. Int.* **78**, 272.

BACHER, R. F. and GOUDSMIT, S. A., 1934. *Atomic Energy States* (International Series in Physics).

BARTLETT, J. H. Jr., 1929. *Phys. Rev.* **34**, 1247.

BATES, D. R., BUCKINGHAM, R. A., MASSEY, H. S. W. and UNWIN, J. J., 1939. *Proc. Roy. Soc.* A, **170**, 322.

BATES, D. R. and MASSEY, H. S. W., 1941. *Proc. Roy. Soc.* A, **177**, 329.

BATES, D. R. and MASSEY, H. S. W., 1943. *Phil. Trans.* A, **239**, 269.

BOYCE, J. C., MENZEL, D. H. and PAYNE, C. H., 1933. *Proc. Nat. Acad. Sci., Wash.*, **19**, 581.

CHAPMAN, S., 1930 a. *Phil. Mag.* **10**, 369.

CHAPMAN, S., 1930 b. *Mem. R. Met. Soc.* **3**, 103.

CHAPMAN, S., 1931. *Proc. Roy. Soc.* A, **132**, 353.

CHAPMAN, S., 1937. *Phil. Mag.* **23**, 657.

CONDON, E. U., 1934. *Astrophys. J.* **79**, 217.

DÉJARDIN, G., 1936. *Rev. Mod. Phys.* **8**, 1.

ELVEY, S. T., SWINGS, P. and LINKE, W., 1940. *Astrophys. J.* **93**, 337.

FRERICHS, R. and CAMPBELL, J. S., 1930. *Phys. Rev.* **36**, 1430.

HAGSTRUM H. D. and TATE, J. T,. 1941. *Phys. Rev.* **59**, 354.

HARTREE, D. R., HARTREE, W. and SWIRLES, B., 1939. *Phil. Trans.* A, **238**, 229.

KENTY, C., 1928. *Phys. Rev.* **32**, 624.

LOZIER, W. W., 1934. *Phys. Rev.* **46**, 268.

MARTYN, D. F. and PULLEY, O. O., 1936. *Proc. Roy. Soc.* A, **154**, 455.

MASSEY, H. S. W. 1938. *Negative Ions* (Cambridge : University Press).

MOHLER, F. L., 1937. *Bur. Stand. J. Res., Wash.*, **19**, 447 and 559.

NICOLET, M., 1939. *Inst. Roy. Météorologique de Belgique*, **11**, 1.

RAYLEIGH, Lord, 1930. *Proc. Roy. Soc.* A, **129**, 458.

SAHA, M. N. and RAI, R. N., 1938. *Proc. Nat. Inst. Sc. India*, **1**, 227.

STEVENSON, A. F., 1932. *Proc. Roy. Soc.* A, **137**, 298.

TÜXEN, O., 1936. *Z. Phys.* **103**, 463.

WULF, O. R. and DEMING, L. S., 1938. *Terr. Mag.* **43**, 282.

YAMANOUCHI, T., INUI, T. and AMEMIYA, A., 1940. *Proc. Phys.-Math. Soc. Japan*, **22**, 84.

YAMANOUCHI, T. and KATANI, M., 1940. *Proc. Phys.-Math. Soc. Japan*, **22**, 60.

PHOTOCHEMICAL PROCESSES IN AN OXYGEN-NITROGEN ATMOSPHERE

By C. H. BAMFORD,
Laboratory of Physical Chemistry, Cambridge

§1. INTRODUCTION.

Applicability of experimental results to photo-reactions in the upper atmosphere

A NUMBER of points concerning the application of photochemical data to processes occurring in the upper atmosphere call for consideration. Firstly, a great deal of the experimental work has been performed, for obvious reasons, at comparatively high pressures. Frequently extrapolation to lower pressures is impossible, through lack of data, or questionable on theoretical grounds. At the same time experimental work at low pressures generally presents much practical difficulty. Secondly, whereas atmospheric processes are necessarily homogeneous, surface reactions generally present a complicating feature in laboratory experiments; in some cases reactions are difficult to observe in the gas phase, owing to the ease with which surface reactions occur. This is, of course, particularly true at low pressures. Thirdly, the number of photochemical reactions which have been studied over reasonably wide ranges of temperature and incident wave-length is unfortunately very small. Finally, a general difficulty is encountered in assessing probabilities of secondary reactions due to the fact that the use, in photochemical reaction schemes, of reaction probabilities derived from thermal data must be regarded with caution, unless (as is hardly ever the case) the energy states of the intermediaries are known.

§2. THE PHOTOCHEMISTRY OF OXYGEN, OZONE AND THE OXIDES OF NITROGEN

(i) *Oxygen*

Oxygen has four regions of absorption. In the infra-red and visible are situated the "atmospheric" absorption bands. These are extremely weak, and are probably due to the forbidden transition $^3\Sigma \rightarrow {}^1\Sigma$ (Badger and Mecke, 1930; Dieke and Babcock, 1927; Babcock, 1930; Mulliken, 1932). The energy absorbed is less than 40 k cal., and thus could not cause dissociation of the molecule, which requires at least 118 k cal. (corresponding to a wave-length of about 2400 A.). No photochemical studies have been made in this region. The Herzberg bands, lying in the quartz ultra-violet region, converge near 2400 A., and are also weak (Herzberg, 1932; Finkelnburg and Steiner, 1932). Warburg (1921) established the formation of ozone from oxygen irradiated by light of wave-length 2530 A. The nature of the primary act at this wave-length is rather uncertain. The energy is just insufficient for dissociation of the

molecule, but since Warburg used high pressures (over 97 kg./cm^2) there is no difficulty in accounting for the experimental results if one assumes formation of excited molecules with a reasonable lifetime, e.g. 10^{-7} sec. (Noyes and Leighton, 1941). Herzberg (1932) has stated that absorption in this region takes the molecule from the ground state ($^3\Sigma_g^-$) to the $^3\Sigma_u^+$ state, which then reacts as follows:

$$O_2(^3\Sigma_u^+) + O_2 \longrightarrow O_3 + O, \qquad \ldots\ldots(1)$$

$$O + O_2 + M \longrightarrow O_3 + M. \qquad \ldots\ldots(2)$$

There is some evidence that absorption at these wave-lengths may be due to O_3 (Wulf, 1928). Still another possibility, pointed out by Rollefson (1939), is that absorption by molecules in vibrational states with $v \geqslant 1$ would permit direct dissociation:

$$O_2 + h\nu \longrightarrow O + O. \qquad \ldots\ldots(3)$$

Warburg's experimental value for the quantum yield of ozone (ϕ) at 2530 A. is 1·14, for an O_2 pressure of 97 kg./cm^2, the value decreasing with increasing O_2 pressure. The above mechanisms would lead to a maximum value of 2 for ϕ. If the upper state is an excited molecule the low experimental figure may be attributed to de-activation, e.g. by collision or radiation, while if the primary act is a dissociation, the high pressures used would be expected to reduce its efficiency owing to the operation of the Franck-Rabinowitch principle (Franck and Rabinowitch, 1934).

Using a shorter wave-length (2070 A.) Warburg (1921) obtained values for ϕ of 2·06 and 1·54, with initial oxygen pressures of 47·5 and 300 kg./cm^2 respectively. The primary act at this wave-length is generally accepted to be

$$O_2 + h\nu \longrightarrow 2O(^3P), \qquad \ldots\ldots(4)$$

this being followed by (2) above. The maximum value of $\phi = 2$ predicted by these equations is thus reached at 47·5 kg./cm^2, while the smaller value at the higher pressure may again be explained in terms of the Franck-Rabinowitch principle. It may be noted that the termolecular collisions necessary for (2) would be relatively frequent at the high pressures used, while the recombination of oxygen atoms by triple collisions would be slow on account of the slow diffusion of the atoms at high pressures.

Warburg also performed experiments in which air was used in place of 95% oxygen, and found that the quantum yield of ozone was reduced only slightly. Thus there is probably no reaction between oxygen atoms and nitrogen molecules which is sufficiently fast under these conditions to compete with (2). This point is of importance in considering photochemical reactions in the atmosphere, and is again referred to below.

The most important absorption bands of oxygen—those forming the Schumann-Runge system—begin near 2000 A. and converge at 1751 A. The maximum intensity lies in the continuum near 1450 A. Flory (1936) states that in the banded region predissociation into two normal oxygen atoms may occur (reaction (4)): for wave-lengths in the continuum not too near the convergence limit one oxygen atom in the ^1D state is probably formed:

$$O_2 + h\nu \longrightarrow O(^3P) + O(^1D). \qquad \ldots\ldots(5)$$

Photochemical studies in this region have been carried out by Vaughan and Noyes (1930), who used oxygen at atmospheric pressure, and radiation which lay practically wholly below the convergence limit, although it was not entirely monochromatic. They found $\phi = 2{\cdot}0 \pm 0{\cdot}3$, a result which, within experimental error, agrees with those of Warburg at longer wave-lengths. Thus Vaughan and Noyes conclude that no difference in behaviour can be ascribed to ^3P and ^1D oxygen atoms.

Using the xenon lines 1470 A. and 1295 A. and oxygen pressures up to about 8 mm. Hg, Groth (1937a) found that $\phi = 1{\cdot}90 \pm 0{\cdot}05$. This figure is in substantial agreement with the values of ϕ at longer wave-lengths. The rather lower value may be explained by practical considerations; the very strong absorption by oxygen at these short wave-lengths causes most of the reactions to occur in a zone near the window through which the light enters, and if reaction between oxygen atoms and ozone is to be avoided the latter must be removed rapidly from this zone. This was difficult in Groth's apparatus. The same author has also studied ozone formation in mixtures of oxygen with nitrogen and with argon. and found that the yield of ozone is reduced only if the amount of oxygen is less than about 4 per cent. The lower yield in these circumstances is attributed to photochemical decomposition of some ozone. This result again points to the absence of any reasonably fast reaction between nitrogen molecules and oxygen atoms, even when the latter are excited to the ^1D state.

It appears to be generally accepted that reaction (2) has no energy of activation (Schumacher, 1930; Groth, 1937b; see also Bonhoeffer and Harteck, 1933). There may be a steric factor which prevents reaction at every triple collision, but there is no general agreement as to the exact magnitude of this. Thus Schumacher (1932) and Beretta and Schumacher (1932) state that reaction occurs after about 10–100 triple collisions when O_2 is the third body. Harteck (1931) gives the number of triple collisions necessary for reaction as 10. Using Kowalsky's data on the upper pressure limit of reaction between phosphorus and oxygen, Semenoff (1935) calculates the probability of (2) to be approximately 3×10^{-3} per triple collision. The data of Melville and Roxburgh (1934), however, from studies of the photochemical oxidation of phosphine, lead to values of the probability as high as 0·5. These latter values are, of course, obtained by very indirect means. It would seem likely that some steric factor is involved, probably of order of magnitude 0·1. Kistiakowsky (1925) has studied the efficiency of other molecules as third bodies; putting that of oxygen equal to unity, the figures for nitrogen and argon are 0·28, 0·22 respectively. Schumacher (1932), from experiments in the ultra-violet region, gives the values as 0·6 and 0·1.

Oxygen absorbs most strongly in the region 1000–670 A., wherein are situated the Hopfield bands. These wavelengths produce ionization (or pre-ionization) of the O_2 molecule, the O_2^+ ion being left in an excited state. For a further description of the Hopfield bands the reader should refer to Chapman and Price (1936).

(ii) *The photolysis of ozone*

Ozone possesses absorption bands in the infra-red, visible and ultra-violet regions. The first of these is without photochemical significance. The visible

absorption begins at about 8000 A., reaches a maximum near 6000 A., and is small at 4500 A. (Colange, 1927; Wulf, 1930). The ultra-violet absorption beginning around 3500 A. has a maximum at about 2500 A. (Dutheil, 1926; Yakovleva and Kondratjew, 1936; Wulf and Melvin, 1931; Kondratjew, 1930; Yakovleva and Kondratjew, 1932). According to Wulf and Melvin (1931) this absorption consists of a band absorption underlying a continuum. The arrangement of bands into a system is not completely successful, and the extrapolated convergence point (2700 A.) is uncertain.

The primary act is the dissociation

$$O_3 + h\nu \longrightarrow O_2 + O. \qquad \ldots\ldots(6)$$

If both products are formed in the ground states ($^3\Sigma$ and ^3P respectively) the energy necessary to effect this reaction is about 24 k cal., corresponding to about 11,800 A., which is thus the longest wave-length capable of producing a dissociation. At shorter wave-lengths the following are possible:

$$O_3 + h\nu \longrightarrow O_2(^1\Sigma) + O(^3P), \qquad \ldots\ldots(6\,a)$$

$$O_3 + h\nu \longrightarrow O_2(^3\Sigma) + O(^1D), \qquad \ldots\ldots(6\,b)$$

$$O_3 + h\nu \longrightarrow O_2(^1\Sigma) + O(^1D). \qquad \ldots\ldots(6\,c)$$

The longest wave-lengths which will provide sufficient energy for these raections are thus approximately 4600 A., 4100 A. and 2700 A. for reactions (6*a*), (6*b*) and (6*c*) respectively. It has been assumed that absorption of wave-lengths in the ultra-violet produces ^1D oxygen atoms (equation (6*b*)) (Schumacher, 1930: Schumacher and Beretta, 1932), but this has not been proved.

A critical survey of the early work on the photolysis of ozone has been given by Schumacher (1930). Kistiakowsky (1925) using red light (6200 A.) has obtained quantum efficiencies of decomposition of about 2 under suitable conditions. The main reactions may be considered to be (6), followed by

$$O + O_3 \longrightarrow 2O_2 + 95 \text{ k cal.}, \qquad \ldots\ldots(7)$$

and also (2). Clearly the occurrence of this latter reaction will mean that the quantum efficiency is dependent upon the total gas pressure, and also upon the nature of any foreign gases present, as is the case in practice. From Kistiakowsky's results, Schumacher (1930) calculates that the probability of (7) at room temperatures is 5×10^{-4}, the oxygen atom being in the normal state. This result is in agreement with the data obtained by Harteck and Kopsch (1931), using oxygen atoms produced in an electrical discharge. These authors find also that the low efficiency of (7) is due to the existence of an appreciable energy of activation for this reaction. In a later paper Beretta and Schumacher (1932) give the energy of activation as 4–6 k cal.

A considerable amount of attention has been paid to the photolysis of ozone in the ultra-violet. The earliest work was that of Warburg (1913); this was followed by investigations of Beretta and Schumacher (1932); Forbes and Heidt (1934); Heidt and Forbes (1934); and Heidt (1935). Forbes and Heidt (1934) used wave-lengths of 2800 A., 2540 A., and 2100 A., and partial ozone pressures from 10–760 mm. Hg. Water vapour was also present. They obtained quantum efficiencies from 1·6–130. Heidt and Forbes (1934), using dry ozone,

obtained quantum efficiencies up to 6·7. It is evident that the simple atom chain of equations (6), (7) and (2) cannot account for these high values, and some form of energy chain is necessary. An energy chain due to imprisonment of radiation is conceivable; alternatively Heidt (1935) suggests the following scheme, which resembles those proposed by Schumacher (1930) and Ritchie (1934):

$$O_3 + h\nu \longrightarrow O_2^* + O, \qquad \ldots\ldots(8\,a)$$

$$O + O_3 \longrightarrow 2O_2^*, \qquad \ldots\ldots(8\,b)$$

$$O + O_2 + M \longrightarrow O_3 + M, \qquad \ldots\ldots(8\,c)$$

$$O \longrightarrow \tfrac{1}{2}O_2 \text{ surface}, \qquad \ldots\ldots(8\,d)$$

$$O_2^* + O_3 \longrightarrow 2O_2 + O, \qquad \ldots\ldots(8\,e)$$

$$O_2^* + M \longrightarrow O_2 + M. \qquad \ldots\ldots(8f)$$

This mechanism, which is a combination of an energy and an atom chain, gives the experimentally determined variation of quantum efficiency with light intensity. Heidt (1935) identifies the state of the excited oxygen molecules in the above reactions as the upper levels for the atmospheric absorption bands ($^1\Sigma_g^+$), but this is based upon spectroscopic arguments which are not convincing. In the presence of water vapour additional reactions have to be postulated, since Forbes and Heidt (1934) found that ϕ is proportional to the square root of the water-vapour concentration. It is possible that water vapour may enter into reaction with oxygen atoms, with formation of OH radicals, which then react further.

(iii) *Nitrous oxide*

This gas possesses two regions of continuous absorption: one is situated at 2200 A. to 1700 A.; the other begins at 1580 A. and extends far down into the ultra-violet (Henry, 1935 a, 35 b; Sen Gupta, 1934, 35; Duncan, 1936). The earliest photochemical studies on the gas were those of Macdonald (1928), who used light in wave-length region 1990–1860 A. Decomposition into nitrogen, oxygen and nitric oxide was observed, the quantum efficiency of decomposition being 3·9 (molecules N_2O per quantum). Macdonald postulated the primary formation of excited N_2O molecules which subsequently enter into reaction with normal molecules; some of his secondary reactions, however, are untenable on energy considerations. The most recent work is that of Noyes (1937), who discusses the primary processes:

$$N_2O + h\nu \longrightarrow N_2 + O(^3P, {}^1D \text{ or } {}^1S), \qquad \ldots\ldots(10)$$

$$N_2O + h\nu \longrightarrow NO + N(^4S \text{ or } {}^2D). \qquad \ldots\ldots(11)$$

Consideration of the energies of the long wave limits of absorption does not provide any adequate spectroscopic grounds on which to base the above reactions. Process (10) may be followed by

$$O + N_2O \longrightarrow 2\,NO, \qquad \ldots\ldots(12)$$

$$O + N_2O \longrightarrow N_2 + O_2, \qquad \ldots\ldots(13)$$

$$O + O + M \longrightarrow O_2 + M. \qquad \ldots\ldots(14)$$

while the following reactions may be expected after the occurrence of (11):

$$N + N_2O \longrightarrow NO + N_2, \quad \ldots\ldots(15)$$

$$N + N + M \longrightarrow N_2 + M. \quad \ldots\ldots(16)$$

The only reactions which lead to the formation of oxygen are (13) and (14), both of which follow the primary process (10). The occurrence of this latter reaction under the experimental conditions holding in Noyes' work thus seems proved. It is more difficult to decide whether (10) represents the whole of the primary process or whether (11) must be included also. Since nitric oxide is a product of photolysis, Noyes points out that if it could be proved that (12) does not occur, it would follow that (11) must take place. In order to obtain information on this point Henrique, Duncan and Noyes (1938) irradiated mixtures of NO_2 and N_2O by light absorbed only by the former—since it is known that photolysis of NO_2 produces oxygen atoms. The authors conclude that these results favour, but do not definitely prove, the occurrence of (11). Further evidence in the same direction is provided by the work of Zabor and Noyes (1940), who irradiated mixtures of nitrous oxide and hydrogen. They report the formation of H_2O and NH_3, the latter presumably arising from interaction of nitrogen atoms formed by (11), with hydrogen. Further, if sufficient hydrogen is added, no nitric oxide is formed. This may be due to reaction between NH_2 radicals and NO:

$$NH_2 + NO \longrightarrow N_2 + H_2O \quad \ldots\ldots(17)$$

(Bamford, 1939).

The general conclusion is, therefore, that the primary process (11) seems probable, but its occurrence cannot be regarded as definitely proved.

(iv) *Nitric oxide*

The absorption spectrum consists of a number of band systems extending from about 2250 A. to shorter wave-lengths (Leifson, 1926; Lambrey, 1929, 30 a, 30 b; Flory and Johnston, 1935). There is no evidence of a continuum down to the limit of observation (~1300 A.). The photolysis has been studied by Macdonald (1928) for wave-lengths in the aluminium spark region (1990–1860 A.), using pressures of 70–630 mm. Hg, and by Flory and Johnston (1935), for wave-lengths 2537–1775 A. and pressures 0·02–7·12 mm. Hg. Macdonald concluded from the discrete nature of the spectrum that the decomposition occurs through the formation of excited NO molecules, which then react with normal NO. Flory and Johnston, however, showed that the rate of decomposition is independent of the pressure of NO, or added inert gas (N_2), a result which would be unlikely if Macdonald's mechanism was the correct one. They further show that at low pressures the effective wave-length is 1832 A. (in the δ band), while above 33 mm. Hg 1849 A. (in the β band) begins to be important. The primary act at low pressures is considered to be

$$NO + h\nu \longrightarrow N + O. \quad \ldots\ldots(18)$$

a process requiring ~122 k cal., corresponding to a wave-length of about 2320 A. Since this occurs in the banded region, it must be a pre-dissociation, probably $^2\Pi \rightarrow {}^2\Sigma$ occurring in the δ band.

Macdonald reported the formation of N_2O during the photolysis of NO, but this is not supported by Flory and Johnston. The latter dismiss the following secondary reactions as energetically improbable:

$$O + NO \longrightarrow O_2 + N - 4{\cdot}6 \text{ k cal.}, \quad \ldots\ldots(19)$$
$$O + N_2 \longrightarrow NO + N - 47{\cdot}4 \text{ k cal.}, \quad \ldots\ldots(20)$$
$$O + N_2 \longrightarrow N_2O - 12{\cdot}6 \text{ k cal.} \quad \ldots\ldots(21)$$

In particular the fact that added nitrogen does not affect the rate of photolysis appears to show that (21) is unlikely (cf. § 2(i)). Flory and Johnston conclude that the whole of the secondary reactions are the wall processes:

$$O + O + M \longrightarrow O_2 + M(+118 \text{ k cal.}), \quad \ldots\ldots(14)$$
$$N + N + M \longrightarrow N_2 + M(+169 \text{ k cal.}), \quad \ldots\ldots(16)$$
$$N + O + M \longrightarrow NO + M(+122 \text{ k cal.}). \quad \ldots\ldots(22)$$

The quantum efficiency of photolysis, according to the mechanism of equations (17), (14), (16) and (21), should lie between 0·5 and 1·0, according to the relative probabilities of the secondary reactions. The value obtained by Macdonald was 0·75.

The existence of pre-dissociation in the β and γ bands of nitric oxide is uncertain.

(v) *Nitrogen peroxide*

The absorption spectrum of NO_2 is well known. Absorption commences at about 6000 A. and is discrete down to 3700 A., where diffuseness sets in. At 2596 A. sharp bands again appear, to be followed at 2459 A. by a second diffuse region (Leifson, 1926; Harris, 1928; Carwile, 1928; Mecke, 1929; Henri, 1930; Lambrey and Corbière, 1934). The photolysis has been studied by Dickinson and Baxter (1928), Norrish (1931) and by Holmes and Daniels (1934). The over-all reaction is

$$2NO_2 + h\nu \longrightarrow 2NO + O_2. \quad \ldots\ldots(23)$$

The diffuseness of the absorption bands below 3700 A. indicates a pre-dissociation in this region:

$$NO_2 + h\nu \longrightarrow NO + O. \quad \ldots\ldots(24)$$

If the O atom is in the ground (^3P) state, any wave-length shorter than 3900 A. is energetically sufficient for dissociation; this presumably occurs in the first diffuse region. To produce ^1D oxygen atoms a wave-length shorter than 2430 A. is required. Thus ^1D atoms may be formed in the second diffuse region, but there is no evidence for this. Pressures in the range 4–230 mm. Hg have been studied, and all observers agree that the quantum efficiency of decomposition increases from zero at 4360 A. to practically 2 at 3130 A. and shorter wave-lengths. In the intermediate region around 4000 A. the sharpness of the absorption spectrum is an indication of the long life of the excited state. Fluorescence at these wave-lengths has been studied by Norrish (1930) and by Baxter (1930). It is clear that at wave-lengths longer than 3700 A., which is the short wave limit of fluorescence, excited molecules are produced in the primary act; some of these fluoresce, others dissociate (if necessary after

acquisition of thermal energy). As would be expected, fluorescence is marked at 4360 A. where decomposition is practically absent.

The main secondary reaction occurring is

$$O + NO_2 \longrightarrow NO + O_2. \qquad \ldots\ldots(25)$$

which probably has a low energy of activation. Thus Spealman and Rodebush (1935) estimate the probability of (25) as 10^{-5} at 40° C., and Henrique, Duncan and Noyes (1938), using this figure, calculate the energy of activation as 4·3 k cal. Thus reactions (24) and (25) can account satisfactorily for the observed quantum efficiency of 2 at short wave-lengths.

Comparatively little is known about N_2O_4. Holmes and Daniels (1934) state that no decomposition follows absorption of 3660 A. or 3130 A., but they report a quantum efficiency of 0·4 at 2650 A. The small yield may be explained if part of the primary act is a decomposition into $NO_2 + NO_2^*$, followed by fluorescence and recombination. Other possible primary acts are:

$$\begin{aligned} N_2O_4 + h\nu &\longrightarrow NO_2 + NO + O, \\ N_2O_4 + h\nu &\longrightarrow N_2O_3 + O, \end{aligned} \qquad \ldots\ldots(26)$$

each of which may lead to a quantum efficiency of 2.

(vi) *Nitrogen pentoxide*

The absorption of this substance has been studied by Urey, Dawsey and Rice (1929a), Dutta and Sen Gupta (1933), and Jones and Wulf (1937), but their observations are not in complete agreement. The results of Jones and Wulf are probably the most reliable; they report a continuum beginning near 3800 A., increasing rapidly at 2800 A., but showing no maximum down to 2400 A. The other observers disagree about the long wave limit of absorption, but there is general agreement about the existence of a continuum below 2800 A.

Using a wave-length of 2800 A. at 0° C., Holmes and Daniels (1934) find quantum efficiencies of decomposition varying from 0·6 at 51·5 mm. Hg to 0·8 at 100 mm. Hg. Similar results were obtained at 2650 A. The over-all reaction is

$$N_2O_5 \longrightarrow 2NO_2 + \tfrac{1}{2}O_2, \qquad \ldots\ldots(27)$$

and the primary act below about 3000 A. may be

$$N_2O_5 + h\nu \longrightarrow (N_2O_4 \rightleftharpoons 2NO_2) + O \qquad \ldots\ldots(28)$$

(Urey, Dawsey and Rice, 1929). The low quantum yield may be due to re-oxidation of NO_2 by O to NO_3. A secondary reaction could be:

$$O + N_2O_5 \longrightarrow N_2O_4 + O. \qquad \ldots\ldots(29)$$

In the presence of NO_2, N_2O_5 undergoes a sensitised decomposition, the mechanism probably being represented by equations (24) and (29), together with (30) below:

$$N_2O_5 + NO \longrightarrow 3NO_2 \qquad \ldots\ldots(30)$$

(Norrish, 1927).

§ 3. GENERAL CONSIDERATION OF PHOTO-REACTIONS IN OXYGEN AND OXYGEN-NITROGEN ATMOSPHERES

(i) *Oxygen atom equilibrium in an oxygen atmosphere*

The primary source of oxygen atoms will be processes which follow light absorption by molecular oxygen. The reaction

$$2O_2 \longrightarrow O_3 + O \qquad \ldots\ldots(1\,a)$$

is about 95 k cal. endothermic, so that the longest wave-length which has sufficient energy to effect it is about 3000 A. This is, in fact, not far removed from the wave-length in the Herzberg bands (2530 A.) which Warburg (1921) found effective in producing ozone. For wave-lengths lying between about 3000 A. and 2500 A. oxygen atom formation probably occurs by reaction (1), this giving place at shorter wave-lengths to the primary dissociation (4). Absorption in the Schumann-Runge bands will probably be most important: as stated earlier, this will lead to the formation of ^3P and ^1D atoms, and possibly atoms in higher states.

Oxygen atoms will also be formed from ozone, by direct photolysis with wave-lengths shorter than 11 800 A., and by collision of ozone molecules with excited (say $^1\Sigma_g^+$) oxygen molecules (equation (8*e*)). This latter reaction appears to be necessary in order to explain the high quantum efficiencies observed during the photolysis of ozone. Nothing is known of the absolute probability of this reaction, nor of that of the deactivation reaction (8*f*), in which M may be an oxygen or an ozone molecule. However, Schumacher (1932) gives for the ratio of the two probabilities (8*f*)/(8*e*) at room temperatures the values 2·0 and 3·5 for $M = O_2$ and $M = O_3$ respectively.

Oxygen atoms will be removed by reaction with oxygen according to equation (2), which has been discussed in § 1 (i). They may also enter into reaction with ozone (equation (7)), or recombine in ternary collisions (equation (14)). The probability of the former reaction has been discussed earlier (§ 1 (ii)), while that of the recombination reaction may be taken as unity for each ternary collision (Groth, 1937a).

The equilibrium of oxygen atoms in an oxygen atmosphere would thus appear to be governed mainly by the above reactions: the probabilities of most of these are known approximately, at least as regards order of magnitude.

(ii) *The oxygen-nitrogen atmosphere*

The work of Warburg (1921) and of Groth (1937a) on the photochemical formation of ozone in mixtures of oxygen and nitrogen point to the conclusion that there is no appreciable reaction between ^3P or ^1D oxygen atoms and nitrogen molecules at ordinary temperatures. This is substantiated by the observation of Flory and Johnston (1935) that the rate of photolysis of nitric oxide (which produces nitrogen and oxygen atoms) is unaffected by addition of nitrogen. Further evidence has been obtained by Steiner (1937), who could detect no N_2O formation from ^1D oxygen atoms and N_2. It is, of course, possible that the rate of reaction is too small to be detectable under the conditions of experiment, while yet being sufficient to account for the N_2O content of the upper atmosphere. Alternatively, oxygen atoms in energy states higher than ^1D may react to form

N_2O. Otherwise it would appear to be impossible to explain the formation of N_2O from neutral oxygen atoms. It is possible that the N_2O in the upper atmosphere arises from reactions involving ions, e.g. O^+, or from interaction between O atoms and metastable N_2 molecules, which are known to be present.

The alternative reaction between O and N_2.

$$N_2 + O \longrightarrow NO + N - 47{\cdot}4 \text{ k cal.}, \qquad \ldots\ldots(20)$$

is even more strongly endothermic than the reaction representing N_2O formation (21), and is clearly improbable at normal temperatures. Actually the reverse reaction is much more probable (see § 3 (iv)).

It is to be expected that the nitrous oxide present in the upper atmosphere will undergo photolysis according to equations (10) and (11), producing oxygen and nitrogen atoms and nitric oxide. This latter will react with oxygen or ozone to give higher oxides of nitrogen. Since photolysis of the oxides of nitrogen does not appear to produce radicals other than oxygen and nitrogen atoms, the essential problem is the elucidation of the reactions of these atoms with the oxides of nitrogen.

(ii *a*) *Reactions between oxygen atoms and oxides of nitrogen*

As stated earlier, the reaction

$$O + NO_2 \longrightarrow NO + O_2 \qquad \ldots\ldots(25)$$

has a probability of 10^{-5} and an energy of activation of about 4·3 k cal. at 40° c., according to Spealman and Rodebush (1935) and Henrique, Duncan and Noyes (1938). These values are obtained from experiments at a single temperature, and their use at temperatures far removed from this is open to criticism.

Two reactions between nitric oxide and oxygen atoms may be envisaged:

$$NO + O \longrightarrow NO_2 + 48{\cdot}7 \text{ k cal.}, \qquad \ldots\ldots(31)$$

$$NO + O \longrightarrow N + O_2 - 4{\cdot}6 \text{ k cal.} \qquad \ldots\ldots(19)$$

The former will presumably need a triple collision (Spealman and Rodebush, 1935); there are no experimental observations on the probabilities of these reactions, except that of Spealman and Rodebush (1935), who state that (31) is slower than (25).

The reactions

$$N_2O + O \longrightarrow 2NO, \qquad \ldots\ldots(12)$$

$$N_2O + O \longrightarrow N_2 + O_2 \qquad \ldots\ldots(13)$$

have been considered by Noyes and his co-workers (Noyes, 1937; Henrique, Duncan and Noyes, 1938) in discussing the photolysis of N_2O. According to Henrique, Duncan and Noyes, the energy of activation of (12) and (13) is >13 k cal. It may be noted that Musgrave and Hinshelwood (1932) consider that (12) proceeds to a small extent at 1000° c.

Of these reactions (25) is supported by a good deal of experimental evidence, and (31) and (12) would appear likely. There is no definite evidence for the occurrence of (19) or (13).

(ii *b*) *Reactions between nitrogen atoms and oxygen and oxides of nitrogen*

The available data on these reactions are even more fragmentary than those for oxygen atoms. Possible reactions are given below, with heats of reaction:

$$N + NO_2 \longrightarrow N_2 + O_2 + 116 \text{ k cal.}, \quad \ldots\ldots(32)$$
$$N + NO_2 \longrightarrow 2NO + 73 \text{ k cal.}, \quad \ldots\ldots(33)$$
$$N + NO \longrightarrow N_2 + O + 47{\cdot}4 \text{ k cal.}, \quad \ldots\ldots(34)$$
$$N + NO + M \longrightarrow N_2O + M + 34{\cdot}8 \text{ k cal.}, \quad \ldots\ldots(35)$$
$$N + N_2O \longrightarrow NO + N_2 + 134 \text{ k cal.}, \quad \ldots\ldots(15)$$
$$N + O_2 \longrightarrow NO + O + 4{\cdot}6 \text{ k cal.}, \quad \ldots\ldots(36)$$
$$N + O_2 + M \longrightarrow NO_2 + M + 53 \text{ k cal.} \quad \ldots\ldots(37)$$

Reactions (34) and (36) are, of course, the reverse of (20) and (19) respectively. Spealman and Rodebush (1935) report that (33) and (34) are "fairly rapid", while (32) is probably much slower. (35) is the reverse of one of the primary acts in the photolysis of N_2O (equation (11)). The work of Flory and Johnston (1935) suggests that this reaction does not occur readily. Reaction (15) appears to be necessary to account for Noyes' results with N_2O. According to Henrique, Duncan and Noyes (1938), the energy of activation of this reaction is <10 k cal. There is no evidence for or against the occurrence of (36) and (37). In addition to the above reactions there are the three possible recombination reactions between oxygen and nitrogen atoms (forming O_2, N_2, NO respectively), each requiring a ternary collision.

It is clear that much laboratory work is necessary before one can perform any quantitative calculations on the nitrogen-atom equilibrium. At the moment the most important reactions would seem to be (33), (34), and (15).

(ii *c*) *Other reactions*

In addition to the above a number of processes not of photochemical origin must be considered. These will only be considered very briefly here; it is hoped to include them in a further report. Firstly, there are the equilibria

$$N_2O_4 \rightleftharpoons 2NO_2. \quad \ldots\ldots(38a)$$
$$2NO_2 \rightleftharpoons 2NO + O_2. \quad \ldots\ldots(38b)$$

Both these have been studied fairly thoroughly; equilibrium constants for the N_2O_4, NO_2 system have been given by Verhoek and Daniels (1931), and for the NO_2, NO, O_2 system by Zeise (1936). It may be noted that the back reaction in (38 *b*) (i.e. the oxidation of NO by O_2) is a termolecular process with a small negative temperature coefficient (Bodenstein, 1932).

Secondly, interaction between ozone and oxides of nitrogen will be important in regions where mixtures of the components are present. It is known that the lower oxides NO and NO_2 are rapidly oxidized to N_2O_5 by ozone (see, e.g., Schumacher and Sprenger, 1928; Gorodetzki, 1939). Processes responsible for the disappearance of N_2O_5 will be (*a*) photolysis, discussed earlier, (*b*) thermal decomposition, (*c*) reaction (29) between N_2O_5 and O atoms and (*d*) reaction between N_2O_5 and NO.

The thermal decomposition of N_2O_5 into $2NO_2 + \frac{1}{2}O_2$ occurs slowly at room temperatures, and becomes appreciably fast at about 40° c. This reaction has

been much studied; values for the unimolecular velocity constant at various temperatures have been given by Daniels and Johnston (1921), White and Tolman (1925), and Hibben (1930). In the presence of ozone, nitrogen pentoxide brings about the decomposition of the ozone, without undergoing any net decomposition itself. The mechanism of this process, according to Schumacher and Sprenger (1929), and Lowry and Seddon (1938), is

$$\begin{aligned} N_2O_5 &\longrightarrow NO_2 + NO_3, \\ NO_2 + O_3 &\longrightarrow NO_3 + O_2, \\ 2NO_3 &\longrightarrow 2NO_2 + O_2, \\ NO_3 + O_3 &\longrightarrow NO_2 + 2O_2. \end{aligned} \qquad \ldots\ldots(39)$$

However, it is unlikely that these processes will be important on account of the slowness of the thermal decomposition of N_2O_5 at the low temperatures of the upper atmosphere.

As regards process (*d*) above, Busse and Daniels (1927) have shown that the reaction (30) occurs at almost every collision.

The oxygen-nitrogen atmosphere can thus present a complicated picture. The state of affairs in the upper atmosphere is not necessarily so complicated; this will be proved to be the case if experimental observation shows that some of the oxides of nitrogen are present only in negligible amounts.

§4. THE PHOTOLYSIS OF WATER AND HYDROGEN PEROXIDE, AND REACTIONS BETWEEN HYDROGEN AND OXYGEN, AND HYDROGEN AND NITROGEN*

(i) *The photolysis of water vapour*

The extinction coefficient of water begins to rise near about 2000 A. (Granath, 1929). Price (1936) found a continuum between 1785–1550 A., with a maximum near 1700 A., while Rathenau (1934) reported a series of broad diffuse bands extending down to <1000 A.

The primary process may be

$$H_2O + h\nu \longrightarrow H + OH, \qquad \ldots\ldots(40)$$

$$H_2O + h\nu \longrightarrow H_2 + O, \qquad \ldots\ldots(41)$$

both of which are energetically possible in the continuum. There seems to be no doubt that H_2O_2 is formed during the photolysis (see, e.g., Kernbaum, 1909; Coehn and Grote, 1912; Tian, 1926). Rodebush and Wahl (1933), using an electrodeless discharge on water vapour, showed that no atomic oxygen is present in the gases leaving the discharge. The afterglow in these experiments is also very small (Rodebush and Wahl, 1933; Stoddart, 1934; Oldenburg, 1935), showing that it is improbable that both H and O are present. Stoddart (1934) found that addition of oxygen to the water vapour increased the afterglow, while Oldenburg (1935, 37) identified OH by its absorption spectrum in the products of the discharge. Thus (40) almost certainly represents the cause of the reaction in the electrodeless discharge, and there is a strong probability that the primary act in photolysis is similar.

* This section is intended to be mainly a summary of reactions. These reactions are not likely to be of major importance, since the hydrogen content of the upper atmosphere is probably very low.

Rodebush, Wende and Campbell (1937) state that H_2O_2 is not formed in this reaction by combination of OH radicals—thus it must arise from other reactions involving H atoms (cf. equations (52) and (53) below).

(ii) *The photolysis of hydrogen peroxide*

Hydrogen peroxide vapour begins to absorb between 3700 A. and 3800 A., with a continuum beginning near 3100 A. (Urey, Dawsey and Rice, 1929b). The products of photolysis are water and oxygen, so that a primary act of the type

$$H_2O_2 + h\nu \longrightarrow H_2 + O_2 \qquad \ldots\ldots(42)$$

is excluded. Urey, Dawsey and Rice (1929), using light of wave-length 2138–2025 A., obtained fluorescence characteristic of OH radicals and concluded that the primary act is the dissociation

$$H_2O_2 + h\nu \longrightarrow 2OH. \qquad \ldots\ldots(43)$$

Further evidence for this was obtained by von Elbe (1932, 33), who irradiated H_2O_2 in the presence of H_2 and CO. These latter were converted into H_2O and CO_2. These results are difficult to explain if the primary act is

$$H_2O_2 + h\nu \longrightarrow H_2O + O, \qquad \ldots\ldots(44)$$

since O atoms react more readily with H_2O_2 than with H_2 or CO (Kistiakowsky, 1930; Harteck and Kopsch, 1931). Assuming (43) to occur, however, this may be followed by

$$OH + H_2 \longrightarrow H_2O + H, \qquad \ldots\ldots(45)$$

$$OH + CO \longrightarrow CO_2 + H. \qquad \ldots\ldots(46)$$

Reaction (46) has been discussed by Bonhoeffer and Haber (1928) and Lavin and Jackson (1931).

As regards the secondary reactions in the photolysis of H_2O_2, the following have been suggested:

$$OH + OH \longrightarrow H_2O + O \qquad \ldots\ldots(47)$$

(Bonhoeffer and Pearson, 1931);

$$OH + OH \longrightarrow H_2 + O_2 \text{ (wall reaction)} \qquad \ldots\ldots(48)$$

(Rodebush, Wende and Campbell, 1937);

$$OH + OH + H_2O_2 \longrightarrow 2H_2O + O_2 \qquad \ldots\ldots(49)$$

(von Elbe, 1933).

Thus it is impossible at present to come to a definite decision as to the exact nature of the secondary processes.

(iii) *Photo-reactions between hydrogen and oxygen*

The reaction

$$O(^3P) + H_2 \longrightarrow OH + H \qquad \ldots\ldots(50)$$

has been studied by Harteck and Kopsch (1931), who report that it has an energy of activation of about 6 k cal., only one in 10^7 collisions being effective at room temperatures. They could identify no H_2O_2 among the products. Kistiakowsky

(1930) and Kistiakowsky and Millington (1931) have studied the interaction of H_2 and O_2 in two wave-length regions, viz. 1854–1862 A., 1719–1725 A. The former would produce excited molecules or normal (^3P) atoms, while the latter would give ^1D atoms. It was found that the quantum yield of ozone was independent of the hydrogen pressure at the longer wave-lengths, but decreased with increasing hydrogen pressure at shorter wave-lengths, H_2O_2 being formed. These results appear to show that ^3P oxygen atoms react with hydrogen very much more slowly than with oxygen, while ^1D oxygen atoms react with hydrogen more readily. Similar results were obtained by Neujmin and Popov (1934), and are in agreement with the theoretical work of Bear and Eyring (1934), who predict energies of activation of 2–3 and 8 k cal. for the reactions involving ^1D and ^3P atoms respectively.

More recently, Smith and Kistiakowsky (1935) reinvestigated the effects of the two wave-length regions. They conclude that no difference between reaction of O atoms for the short and long wave-length regions is to be drawn, and they give the reactions

$$O + O_2 + M \longrightarrow O_3 + M, \quad \ldots\ldots(51)$$

$$O + H_2 \longrightarrow OH + H, \quad \ldots\ldots(50)$$

$$H + O_2 + M \longrightarrow HO_2 + M, \quad \ldots\ldots(52)$$

$$2OH \longrightarrow H_2O_2, \quad \ldots\ldots(53)$$

$$2HO_2 \longrightarrow H_2O_2 + O_2, \quad \ldots\ldots(54)$$

$$OH + HO_2 \longrightarrow H_2O + O_2, \quad \ldots\ldots(55)$$

and

$$O + H_2 + O_2 \longrightarrow H_2O + O_2. \quad \ldots\ldots(56)$$

This work is in disagreement with the earlier in that H_2O_2 was found at longer wave-lengths. The yield of H_2O_2 was too high to allow (50) to have a considerable energy of activation if (51) has very little, and the authors conclude that probably both are three-body processes, (50) being actually

$$O + O_2 + H_2 \longrightarrow OH + HO_2. \quad \ldots\ldots(57)$$

However, objection can be raised to the above mechanism in that there is no evidence for (53). Thus Bonhoeffer and Pearson (1931) found that the emission spectrum of OH was obtained when a discharge was passed through water vapour, but no H_2O_2 could be isolated.

With regard to the attack of hydrogen atoms on oxygen molecules, it is agreed that, at any rate at two temperatures, the first step is the formation of HO_2. Further, it is generally agreed that this needs a triple collision (reaction (52)) (Cook and Bates, 1935; Bodenstein and Schenk, 1933; Kimball, 1937). Bodenstein and Schenk (1933) concluded that the reaction requires an energy of activation of $\sim$1·5 k cal. Reactions which follow the formation of HO_2 may be (54) above or (58) below

$$HO_2 + H_2 \longrightarrow H_2O_2 + H \quad \ldots\ldots(58)$$

(Bates, 1933 a, 33 b, 33 c). This latter reaction would have an energy of activation of at least 19 k cal., and so would not proceed appreciably at low temperatures. In connection with this reaction see also Bodenstein and Schenk (1933) and Kornfeld (1937).

(iv) *Reactions between hydrogen and nitrogen*

It seems probable that hydrogen atoms are without action on nitrogen molecules and that there is no reaction between excited nitrogen molecules and molecular hydrogen (Lewis, 1928). Both reactants must be excited for reaction to occur. However, Noyes (1925, 30) has found that NH_3 is formed when mixtures of N_2 and H_2 and mercury vapour at pressures of one atmosphere and higher and at the temperature of boiling mercury are irradiated by mercury resonance radiation. Since $^3P_1$ Hg does not excite N_2, this result is surprising; it appears necessary to postulate the dissociation of nitrogen molecules by (metastable) $^3P_0$ Hg, two of the latter dissociating one molecule of N_2. Actually there is other evidence for a mechanism of this kind (Gaviola, 1928, 29; Beutler and Rabinowitch, 1930). Formation of ammonia may then follow the reaction

$$N + H + M \longrightarrow NH + M, \qquad \ldots\ldots (59)$$

followed by further reactions between NH and H or NH and H_2.

REFERENCES

Babcock, H. D., 1930. *Phys. Rev.* **35**, 125.
Badger, R. M. and Mecke, R., 1930. *Z. Phys.* **60**, 59.
Bamford, C. H., 1939. *Trans. Faraday Soc.* **35**, 568.
Bates, J. R., 1933 a. *J. Amer. Chem. Soc.* **55**, 426.
Bates, J. R., 1933 b. *Proc. Nat. Acad. Sci., Wash.*, **19**, 81.
Bates, J. R., 1933 c. *J. Chem. Phys.* **1**, 457.
Bates, J. R., 1935. *Chem. Rev.* **17**, 406.
Baxter, W. P., 1930. *J. Amer. Chem. Soc.* **52**, 3920.
Bear, R. S. and Eyring, H., 1934. *J. Amer. Chem. Soc.* **56**, 2020.
Beretta, U. and Schumacher, H. J., 1932. *Z. phys. Chem.* B, **17**, 417.
Beutler, H. and Rabinowitch, E., 1930. *Z. phys. Chem.* B, **6**, 233.
Bodenstein, M., 1922. *Z. phys. Chem.* **100**, 68.
Bodenstein, M. and Schenk, P. W., 1933. *Z. phys. Chem.* B, **20**, 420.
Bonhoeffer, K. F. and Haber, F., 1928. *Z. phys. Chem.* A, **137**, 263.
Bonhoeffer, K. F. and Harteck, P., 1933. *Grundlagen der Photochemie* (Steinkopff), 216.
Bonhoeffer, K. F. and Pearson, T. G., 1931. *Z. phys. Chem.* B, **14**, 1.
Busse, W. F. and Daniels, F., 1927. *J. Amer. Chem. Soc.* **49**, 1257.
Carwile, L. C. K., 1928. *Astrophys. J.* **67**, 84.
Chapman, S. and Price, W. C., 1936. *Rep. Prog. Phys.* **3**, 58.
Coehn, A. and Grote, G., 1912. *Nernst Festschrift* (Halle : Wilhelm Knapp), 136.
Colange, G., 1927. *J. Phys. Radium*, **8**, 256.
Cook, G. A. and Bates, J. R., 1935. *J. Amer. Chem. Soc.* **57**, 1775.
Daniels, F. and Johnston, E. H., 1921. *J. Amer. Chem. Soc.* **43**, 53.
Dickinson, R. G. and Baxter, W. P., 1928. *J. Amer. Chem. Soc.* **50**, 774.
Dieke, G. H. and Babcock, H. D., 1927. *Proc. Nat. Aacd. Sci., Wash.*, **13**, 670.
Duncan, A. B. F., 1936. *J. Chem. Phys.* **4**, 638.
Dutheil, J., 1926. *J. Phys. Radium*, **7**, 415.
Dutta, A. K. and Sen Gupta, P. K., 1933. *Proc. Roy. Soc.* A, **139**, 397
Elbe, G. von, 1932. *J. Amer. Chem. Soc.* **54**, 821.
Elbe, G. von, 1933. *J. Amer. Chem. Soc.* **55**, 62.
Finkelnburg, W. and Steiner, W., 1932. *Z. Phys.* **79**, 69.
Flory, P. J., 1936. *J. Chem. Phys.* **4**, 23.
Flory, P. J. and Johnston, H. L., 1935. *J. Amer. Chem. Soc.* **57**, 2641.
Forbes, G. S. and Heidt, L. J., 1934. *J. Amer. Chem. Soc.* **56**, 1671.
Franck, J. and Rabinowitch, E., 1934. *Trans. Faraday Soc.* **30**, 120.
Gaviola, E., 1928. *Nature, Lond.*, **122**, 313.
Gaviola, E., 1929. *Phys. Rev.* **34**, 1373.

GORODETZKI, G. A., 1939. *J. Appl. Chem., U.S.S.R.*, **12**, 1637.
GRANATH, L. P., 1929. *Phys. Rev.* **34**, 1045.
GROTH, W., 1937 a. *Z. phys. Chem.* B, **37**, 307.
GROTH, W., 1937 b. *Z. phys. Chem.* B, **37**, 315.
HARRIS, L., 1928. *Proc. Nat. Acad. Sci., Wash.*, **14**, 690.
HARTECK, P., 1931. *Naturwissenschaften*, **19**, 858.
HARTECK, P. and KOPSCH, U., 1931. *Z. phys. Chem.* B, **12**, 327.
HEIDT, L. J., 1935. *J. Amer. Chem. Soc.* **57**, 1710.
HEIDT, L. J. and FORBES, G. S., 1934. *J. Amer. Chem. Soc.* **56**, 2365.
HENRI, V., 1930. *Nature, Lond.*, **125**, 202.
HENRIQUE, F. C., DUNCAN, A. B. F. and NOYES, W. A., 1938. *J. Chem. Phys.* **6**, 518.
HENRY, L., 1935 a. *C.R. Acad. Sci., Paris*, **200**, 656.
HENRY, L., 1935 b. *J. Chim. Phys.* **32**, 437.
HERZBERG, G., 1932. *Naturwissenschaften*, **20**, 577.
HIBBEN, J. H., 1930. *J. Phys. Chem.* **34**, 1387.
HOLMES, H. H. and DANIELS, F., 1934. *J. Amer. Chem. Soc.* **56**, 630.
JONES, E. J. and WULF, O. R., 1937. *J. Chem. Phys.* **5**, 873.
KERNBAUM, M., 1909. *C. R. Acad. Sci., Paris*, **149**, 273.
KIMBALL, G. E., 1937. *J. Chem. Phys.* **5**, 310.
KISTIAKOWSKY, G. B., 1925. *Z. phys. Chem.* **117**, 337.
KISTIAKOWSKY, G. B., 1930. *J. Amer. Chem. Soc.* **52**, 1868.
KISTIAKOWSKY, G. B. and MILLINGTON, P. E., 1931. *Z. phys. Chem.* B, **13**, 155.
KONDRATJEW, V., 1930. *Z. phys. Chem.* B, **7**, 70.
KORNFELD, G., 1937. *Z. phys. Chem.* B, **35**, 236.
LAMBREY, M., 1929. *C.R. Acad. Sci., Paris*, **189**, 574.
LAMBREY, M., 1930 a. *C.R. Acad. Sci., Paris*, **190**, 670.
LAMBREY, M., 1930 b. *Ann. Phys., Paris* [x], **14**, 95.
LAMBREY, M. and CORBIÈRE, J., 1934. *C.R. Acad. Sci., Paris*, **201**, 1334.
LAVIN, G. I. and JACKSON, W. F., 1931. *J. Amer. Chem. Soc.* **53**, 383 and 3189.
LEIPSON, S., 1926. *Astrophys. J.* **63**, 73.
LEWIS, B., 1928. *J. Amer. Chem. Soc.* **50**, 27.
LOWRY, T. M. and SEDDON, R. V., 1938. *J. Chem. Soc.* 626.
MACDONALD, J. Y., 1928. *J. Chem. Soc.* 1.
MECKE, R., 1929. *Naturwissenschaften*, **17**, 996.
MELVILLE, H. W. and ROXBURGH, H. L., 1934. *J. Chem. Phys.* **2**, 747.
MULLIKEN, R. S., 1932. *Rev. Mod. Phys.* **4**, 55.
MUSGRAVE, F. F. and HINSHELWOOD, C. N., 1932. *Proc. Roy. Soc.* A, **135**, 23.
NEUJMIN, H. and POPOV, B., 1934. *Z. phys. Chem.* B, **27**, 15.
NORRISH, R. G. W., 1927. *Nature, Lond.*, **119**, 123.
NORRISH, R. G. W., 1931. *J. Chem. Soc.* 1158, 1604 and 1611.
NOYES, W. A., 1925. *J. Amer. Chem. Soc.* **47**, 1003.
NOYES, W. A., 1930. *J. Amer. Chem. Soc.* **52**, 2418.
NOYES, W. A., 1937. *J. Chem. Phys.* **5**, 807.
NOYES, W. A. and LEIGHTON, P. A., 1941. *The Photochemistry of Gases* (Reinhold), 247.
OLDENBERG, O., 1935. *J. Chem. Phys.* **3**, 266.
OLDENBERG, O., 1937. *J. Phys. Chem.* **41**, 293.
PRICE, W. C., 1936. *J. Chem. Phys.* **4**, 147.
RATHENAU, G., 1934. *Z. Phys.* **87**, 32.
RITCHIE, M., 1934. *Proc. Roy. Soc.* A, **146**, 848.
RODEBUSH, W. H. and WAHL, M. H., 1933. *J. Chem. Phys.* **1**, 696.
RODEBUSH, W. H., WENDEL, C. W. J. and CAMPBELL, R. W., 1937. *J. Amer. Chem. Soc.* **59**, 1924.
ROLLEFSON, G. K. and BURTON, M., 1939. *Photochemistry* (Prentice-Hall), 328.
SCHUMACHER, H. J., 1930. *J. Amer. Chem. Soc.* **52**, 2377.
SCHUMACHER, H. J., 1932. *Z. phys. Chem.* B, **17**, 405.
SCHUMACHER, H. J. and SPRENGER, G., 1928. *Z. phys. Chem.* **136**, 77 and **140**, 281.
SCHUMACHER, H. J. and SPRENGER, G., 1929. *Z. angew. Chem.* **42**, 627.
SEMENOFF, N., 1935. *Chemical Kinetics and Chain Reactions* (Oxford : Clarendon Press), 181.

SEN GUPTA, P. K., 1934. *Bull. Acad. Sci. United Provinces Agra Oudh (India)*, **3**, 197.
SEN GUPTA, P. K., 1935. *Nature, Lond.*, **136**, 513.
SMITH, H. A. and KISTIAKOWSKY, G. B., 1935. *J. Amer. Chem. Soc.* **57**, 835.
SPEALMAN, M. L. and RODEBUSH, W. H., 1935. *J. Amer. Chem. Soc.* **57**, 1474.
STEINER, A. B., 1937. *Thesis*, Stanford University.
STODDART, E. M., 1934. *Phil. Mag.* [2] **18**, 410.
TIAN, A., 1916. *Ann. Phys., Paris* [ix], **5**, 248.
UREY, H. C., DAWSEY, L. H. and RICE, F. O., 1929 a. *J. Amer. Chem. Soc.* **51**, 3190.
UREY, H. C., DAWSEY, L. H. and RICE, F. O., 1929 b. *J. Amer. Chem. Soc.* **51**, 1371.
VAUGHAN, W. E. and NOYES, W. A., 1930. *J. Amer. Chem. Soc.* **52**, 559.
VERHOEK, F. H. and DANIELS, F., 1931. *J. Amer. Chem. Soc.* **53**, 1250.
WARBURG, E., 1913. *Sitzb. kgl. preuss. Akad.* 644.
WARBURG, E., 1921. *Z. Elektrochem.* **27**, 133.
WHITE, E. C. and TOLMAN, R. C., 1925. *J. Amer. Chem. Soc.* **47**, 1240.
WULF, O. R., 1928. *Proc. Nat. Acad. Sci., Wash.*, **14**, 356 and 603.
WULF, O. R., 1930. *Proc. Nat. Acad. Sci., Wash.*, **16**, 508.
WULF, O. R. and MELVIN, E. H., 1931. *Phys. Rev.* **38**, 333.
YAKOVLEVA, A. V. and KONDRATJEW, V., 1932. *Phys. Z. Sowjet.* **1**, 472.
YAKOVLEVA, A. V. and KONDRATJEW, V., 1936. *Phys. Z. Sowjet.* **9**, 106.
ZABOR, J. W. and NOYES, W. A., 1940. *J. Amer. Chem. Soc.* **62**, 1975.
ZEISE, H., 1936. *Z. Elektrochem.* **42**, 785.

THE PHOTOCHEMISTRY OF ATMOSPHERIC OXYGEN

By SYDNEY CHAPMAN, F.R.S.,
Imperial College, London

§1. INTRODUCTION

Oxygen in the atmosphere exists mainly in the diatomic form O_2, but partly also in the atomic and triatomic forms O and O_3 (ozone). The presence of ozone is proved by direct chemical analysis of the air near the ground, and by spectroscopic measurements made at the ground and at higher levels up to 30 km. The presence of atomic oxygen is indicated by the atmospheric emission, observed at night, of certain lines in the O I spectrum; the level of emission is likely to be about 100 km. or more, but is not actually known.

The literature relating to atmospheric ozone has been summarized, with bibliographies, by Götz (1931, 38).

The present *Report* outlines the attempts made to account for the presence of the abnormal forms of atmospheric oxygen, O and O_3. The following are the principal facts to be explained regarding the ozone.

The total amount of atmospheric ozone, averaged over the whole earth, is about 2·5 atmo-mm.* The annual mean value increases, however, with the latitude, from about 2·2 atmo-mm. at the equator to 3 at 60° latitude, and to still higher values nearer the poles. There is a seasonal variation (except at the equator), increasing with the latitude; at Lerwick (60° N.) the O_3 has a minimum of 2·5 atmo-mm. in the autumn and a maximum of 3·5 in the spring.

In temperate latitudes, at least, the amount of O_3 above any observing station is found to vary considerably from day to day, and sometimes even from hour to hour; these temporal changes do not necessarily imply rapid formation and destruction of the ozone; they may be due to the passage of an irregular horizontal distribution of O_3 in the air stream as this flows over the station. When thunderstorms pass overhead the O_3 content has been found to increase considerably; e.g. on two thundery days in July 1941, Dobson at Oxford observed a rise of about 2 atmo-mm. (from 2·7 to 4·7 atmo-mm.), lasting from one to three or four hours. The day-to-day variations are correlated with the weather changes. The annual mean ozone content seems nearly constant throughout the sunspot cycle (Götz, 1931, p. 215).

* The total amount of any atmospheric constituent, in any complete vertical column, may be indicated by giving the thickness of the layer formed if the constituent were separated out and spread over the base of the column at S.T.P. So expressed, the amount may be said to be given in atmo-units of thickness (e.g. atmo-cm., atmo-mm., or atmo-km.).

The amount of the constituent in an incomplete vertical column, between any two assigned heights, can be specified in the same way; and the *density* at any particular height can be expressed in atmo-units of thickness per km. (say) at that height.

The vertical distribution of the ozone up to about 30 km. has been determined by three different methods, which give reasonably accordant results (Götz, 1938, pp. 274–293). The value of the maximum ozone *density* depends on the total ozone *content*; in Europe, when the content is 2 atmo-mm., the maximum density is 0·1 atmo-mm. per km.; it increases to 0·15 when the content is 3 atmo-mm. The maximum density occurs at from 20 to 25 km. height; before 1931 much higher estimates—40 to 50 km. were given, based on less accurate methods. The density decreases to about 0·05 atmo-mm. per km. near the ground (in unpolluted air); this value does not vary much with the total ozone content, showing that the main ozone changes occur in the stratosphere. The decrease of the ozone density from its level of maximum is supposed to be more rapid in the upward than in the downward direction.

Owing to the upward decrease of air density, the ozone *concentration* (or volume proportion) reaches its maximum, about 7×10^{-6}, at about 40 km. height.

§ 2. THEORETICAL DISCUSSIONS

Theories of the ozone content of the atmosphere have been given by Chapman (1930a–30c), Mecke (1931a, 31b), Wulf (1932–35) and Wulf and Deming (1936a–37). Chapman and Mecke wrote at the time when the level of maximum ozone density was supposed to be at 40 to 50 km.; they have not published revisions of their work.

These discussions greatly idealize the problem. They all treat the atmosphere as static, and thus exclude from their consideration the horizontal and vertical transfer of ozone by winds and turbulence, and the consequent day-to-day changes of ozone content. They all give a purely photochemical treatment of the problem, involving only O_2, O and O_3; thus they ignore any reaction with nitrogen or other atmospheric constituents, and any ozone changes due to causes other than sunlight, such as thunderstorms, or corpuscular radiation from the sun or elsewhere. They all neglect ordinary diffusion; Chapman (1930a, p. 105) showed that this is unimportant from 40 km. downward, so that if ozone is formed only in the stratosphere (§ 3) its occurrence in the troposphere must be due to downward transfer by the convective turbulence almost always present.

Chapman (1930a) first considered the changing balance between the formation and destruction of ozone in the main layer of greatest density; he considered the regular daily and annual O_3 variations (see also Chapman, 1930b) and the latitude-dependence of the annual mean O_3 content. Later (1930c) he considered the vertical distribution of O_3 and O up to the limit of the atmosphere, showing that above some level, perhaps at 100 km. or so, the oxygen must exist mainly in the atomic form, mixed with O_2 and a far smaller proportion of O_3.

Mecke (1931a, 31b) considered the height distribution of the ozone, supposing it to be in equilibrium under constant solar radiation. He did not discuss the rate at which the equilibrium would vary seasonally or by day and night. His main result was

$$n_3 = (n_3)_{max} (p/p_m)^2 \, e^{2(1-p/p_m)}, \qquad \ldots\ldots(1)$$

where n_3 denotes the number of O_3 molecules per c.c., at the level where the

pressure is p, and $(n_3)_{max}$ denotes the maximum value of n_3, occurring where $p = p_m$. The value of p_m was taken to be given by the height (then supposed to be 50 km.) of maximum n_3; this height was supposed to be accordant with the absorption coefficient (0·020 $cm.^{-1}$) for the ultraviolet light (at about λ 1850) to which ozone formation was attributed. The value of $(n_3)_{max}$ in (1) was inferred by equating the integral of n_3 with respect to height, according to (1), to the known ozone content. The formula (1) is based on qualitative or relative photochemical arguments, and does not involve any measured photochemical reaction constants. The same remarks apply to Chapman's discussion, though he inferred the values of certain reaction constants from the observed regular variations of atmospheric ozone.

Wulf and Deming (1936 a–37) calculated the ozone equilibrium that would be set up in the earth's atmosphere, at all heights, if the sun remained stationary vertically above the point considered. Using observed absorption coefficients for sunlight, and assuming that the sun's spectrum is that of a black body at 6000° K., they calculated n_3 at all heights up to about 60 km. Their results depended on a single unknown, a ratio of two reaction constants (k_{13}/k_{12}, § 4); for this they adopted a value that led to the observed total ozone content. They also calculated the rate at which half the ozone would be restored, if it were all suddenly destroyed; they estimated that the time required would be only a few minutes at altitudes well above the level of maximum n_3, and would increase downwards, to be several days at levels below the maximum.

§ 3. PHOTOLYSIS BY THE ABSORPTION OF SUNLIGHT

Ozone can be formed from O_2 by the absorption of sunlight within the ultraviolet range $\lambda < 2530$. In part of this range the O_2 molecule is directly dissociated, in the longer-wave part it is merely excited, and ozone-formation occurs only when the excited molecule, before it can radiate its excess energy, collides with another O_2 molecule (see the *Report* by Bamford, § 2); the O atoms set free may combine with O_2 molecules to form O_3.

If these radiations were (i) emitted with known intensity from the sun, and were (ii) absorbed in the earth's atmosphere only by oxygen, and (iii) if $\alpha_2(\lambda)$, the O_2 absorption coefficient at wavelength λ, were known, it would be possible to determine the number of absorbed quanta of each wavelength, at each oxygen mass-level in the atmosphere; "mass-level" here signifies the level above which the oxygen, in a vertical column of 1 sq. cm. cross-section, has a known mass. The *height* corresponding to this level need not be known, and will depend on the vertical distribution of atmospheric temperature and density.

Unfortunately none of the above three conditions is satisfied.

(i) The intensity of the sun's ultraviolet spectrum is unknown; in the ozone discussions by Chapman, and by Wulf and Deming, it was supposed to correspond to a black body of temperature 6000° K. It may be somewhat greater than this, but the discrepancy is probably not serious until we reach much shorter (ionizing) wavelengths than those of interest here (see the *Report* by Hunter).

(ii) The radiation between λ2530 and λ1800 is absorbed by O_2 as well as by O_3, and the O_3 absorption coefficient $\alpha_3(\lambda)$ here much exceeds $\alpha_2(\lambda)$. More

over, (iii) the values of both $\alpha_2(\lambda)$ and $\alpha_3(\lambda)$ are seriously uncertain in parts of this range of λ. Even if $\alpha(\lambda)$ for each gas were fully known, the fact that there is absorption by both O_3 and O_2 complicates the calculation of n_3, as attempted by Wulf and Deming, because the initially unknown value of n_3 partly determines the absorption on which the ozone formation depends.

The same difficulty arises in connection with the photolysis of O_3, by radiation for which $\lambda<11{,}800$. Within this range the absorption by O_3 occurs mainly in the ultraviolet region ($\lambda<3800$), but there is also slight absorption in the Chappuis bands ($4400<\lambda<8000$) of the visible spectrum (Wulf and Deming, 1936 b). Diatomic oxygen absorbs ultraviolet light not only, as already mentioned, in the range $\lambda<2530$ associated with ozone formation, but also for $\lambda>2530$; such absorption weakens the photolysis of O_3 without dissociating the O_2. Moreover, O_2 absorbs (though weakly) in the electronic "atmospheric" oxygen band ($^3\Sigma - {}^1\Sigma$) around λ 7600; this absorption will slightly reduce the direct photolysis of O_3 by Chappuis absorption, but according to Mecke (1931 c, p. 364) the metastable O_2 molecules ($^1\Sigma$) so produced may assist in the dissociation of ozone; such metastable molecules may also be produced by the ultraviolet photolysis of O_3 (Mecke, 1931 a, p. 396).

Another difficulty of a different kind is that the absorption coefficient $\alpha_3(\lambda)$ for O_3 in the ultraviolet region depends materially on the temperature T (Wulf and Melvin, 1931; Vassy, 1936); in the atmosphere T is known up to 30 km., above which its value is uncertain, though it is believed to rise to values of 300° c. or even 400° c. at heights of 50 or 60 km.

Wulf and Deming (1937), in their third paper on the ozone equilibrium, considered three alternative distributions of $\alpha_3(\lambda)$ (differing only in the region λ 2100 to λ 1700); they calculated the absorption of light by O_2 and O_3 in several narrow bands of wavelength, in several intervals of height (each of 5 or 10 km.), starting from high levels and working downwards; in each height-interval the equilibrium value of n_3 was calculated, by a process of successive approximation. The intensity of the light passing into the next underlying layer could then be determined. The density of O_2 at each level could be considered as known.

Their three calculations gave the following as the (alternative) height-intervals in which n_3 attained its maximum: 30–35 km.; 35–40 km.; 40–45 km. All these intervals are above the observed height of maximum n_3. Wulf and Deming suggest two possible explanations of the discrepancy: it may be due to the strong temperature-dependence of the ratio k_{12}/k_{13} on which their results depend (§ 4), or the theory of photochemical equilibrium may apply only to the upper parts of the ozone region.* In the latter case, the upper part might shield the main lower part of the ozone from photolysis by ultraviolet light, most of this having already been absorbed above: the ozone below 25 or 30 km. would on this view not be formed *in situ*, but be carried down from the upper levels by convection.

One inference to be drawn from the discussion by Wulf and Deming is that the treatment by Mecke is altogether too simple to allow much weight to be given to his formula (1).

* In this case their estimates of k_{12}/k_{13} would cease to be appropriate.

§4. THE PHOTOCHEMICAL REACTIONS

The primary process of ozone formation is taken to be the dissociation of oxygen :

$$O_2 + h\nu = 2O; \qquad \ldots\ldots(2)$$

some contribution will also be made by the double process

$$O_2(^3\Sigma) + h\nu = O_2(^3\Sigma_u^+), \qquad O_2(^3\Sigma_u^+) + O_2 = O_3 + O; \qquad \ldots\ldots(3)$$

but in the main part of the ozone layer the radiation involved in (3) will be small, because most of the radiation of the wavelength involved is absorbed by the ozone at higher levels.

The photolysis of O_3 likewise produces O atoms :

$$O_3 + h\nu = O_2 + O. \qquad \ldots\ldots(4)$$

The O atoms produced in any of these three ways may, by three-body collisions, either re-unite to form O_2, or unite with O_2 to form O_3:

$$2O + M = O_2 + M, \qquad \ldots\ldots(5)$$

$$O + O_2 + M = O_3 + M; \qquad \ldots\ldots(6)$$

here M signifies any third molecule, which in the atmosphere will most often be N_2 or O_2.

The ozone formed by the reactions (3) or (6) may combine with O atoms or with other O_3 molecules in two-body collisions :

$$O + O_3 = 2O_2, \qquad \ldots\ldots(7)$$

$$2O_3 = 3O_2. \qquad \ldots\ldots(8)$$

The last reaction (the thermal decomposition of ozone) takes place very slowly at ordinary temperatures, and still more slowly in the lower stratosphere, where $T = 220°$ K.; hence it has been left out of account in all the photochemical discussions of atmospheric ozone.

Let n_2, n_3, n_1 and n denote the respective numbers of the atoms or molecules O_2, O_3, O and M per c.c.; they are all functions of height; n_2 and n may be considered known throughout the ozone region. The rates of change of n_1, n_3 and n_2 are given by the following equations :

$$dn_1/dt = 2Q_2 + Q_3 - k_{12}n_1n_2n - 2k_{11}n_1^2n - k_{13}n_1n_3, \qquad \ldots\ldots (9)$$

$$dn_3/dt = k_{12}n_1n_2n - Q_3 - k_{13}n_1n_3, \qquad \ldots\ldots(10)$$

$$dn_2/dt = -Q_2 - k_{12}n_1n_2n + k_{11}n_1^2n + 2k_{13}n_1n_3. \qquad \ldots\ldots(11)$$

Here Q_2 denotes the number of O_2 molecules dissociated per c c. per second, and Q_3 the corresponding number of O_3 molecules dissociated.

§5. THE PHOTOCHEMICAL EQUILIBRIUM (WULF AND DEMING)

Wulf and Deming calculated Q_2, Q_3 and n_3 over a range of height for *steady vertically incident* sunlight. They determined the corresponding equilibrium value of n_3, assuming that the reaction (5) was negligible in the ozone layer.

Hence, writing $k_{11}=0$ and $dn_1/dt=0$, $dn_3/dt=0$ in (9) and (10), they obtained the results

$$n_1=(2Q_2+Q_3)/(k_{12}n_2n+k_{13}n_3), \quad \ldots\ldots(12)$$

$$n_1=Q_3/(k_{12}nn_2-k_{13}n_3), \quad \ldots\ldots(13)$$

and hence, also,

$$n_3=(k_{12}/k_{13})nn_2Q_3/(Q_2+Q_3). \quad \ldots\ldots(14)$$

In calculating Q_2 and Q_3, Wulf and Deming took the density distribution of the air and of oxygen to be that given by Chapman and Milne (1920), in which diffuse separation of the main constituents occurs above 20 km., and $T=219°$ K. everywhere above the tropopause. This distribution does not, in fact, apply to the actual atmosphere, but at the levels here considered (up to 60 or 70 km.) the resulting error is probably not serious for the theory of the ozone region.

Their calculations of Q_2 and Q_3 (§ 3) involved the value of n_3, which, however, is given only when Q_3 is known; hence a process of successive approximation was used to obtain a consistent set of values of Q_3 and n_3 (Wulf and Deming, 1936a, p. 308; 1937, p. 197). In these calculations it was necessary also to know (or adopt) the value of the reaction ratio k_{12}/k_{13}; in their first paper they made a preliminary estimate that its value would be of order 10^{-19}, using rough kinetic-theory considerations as to the relative frequency of three-body and two-body collisions, and assuming (on the basis of certain laboratory experiments) that at a three-body collision the reaction (6) is 10^4 times as probable as (7) is at a two-body collision. In their calculations they chose (by trial) a value of k_{12}/k_{13}, which would give the observed total ozone content; in their first paper they thus took k_{12}/k_{13} as 10^{-21}; in their third paper the values they found appropriate were 10^{-19}, 10^{-19} and 10^{-18} in their three alternative calculations (A, B, C), which differed in the adopted values of $\alpha_3(\lambda)$.

§ 6. THE REGULAR DAILY OZONE VARIATION

The special value of the discussions by Wulf and Deming lies in their calculation and use of Q_2 and Q_3, based on the best available data for $\alpha_2(\lambda)$ and $\alpha_3(\lambda)$: but they considered only the steady state under constant vertically incident sunlight. The bearing of their results on the actual ozone distribution consequently cannot be judged without further theoretical study. Their calculations of Q_2 and Q_3 will also need revision when the present uncertainties are resolved as to the values of $\alpha_2(\lambda)$ and $\alpha_3(\lambda)$ in certain ranges of λ.

Chapman (1930a) discussed the regular daily and seasonal variations of ozone near its level of maximum density. No regular daily ozone variation has been observed; he therefore inferred that its range will not exceed one-tenth of the whole ozone present. On the basis of this estimate, and of the known seasonal variation, and of the ratio Q_3/Q_2 at the level of maximum ozone, he estimated the values of k_{12}, k_{13} and k_{11}; he took Q_3/Q_2 to be 600; Wulf and Deming, in their case C, calculated the value of Q_3/Q_2 at 27·5 km. to be 820. By simple arguments Chapman showed that the ratios $k_{12}n_2n/k_{13}n_3$ and (at noon) $k_{12}n_2n_1/k_{11}n_1^2$ are at least as great as Q_3/Q_2. He showed also that in consequence (at the level of maximum ozone)

$$n_1=Q_3/k_{12}n_2n \quad \ldots\ldots(15)$$

is very nearly true; this agrees with (12) and (13) if the relatively small terms in denominators or numerator are ignored. He inferred also that at this level the time required for n_1 to approach its equilibrium value is brief (only a few minutes), and that n_1 has a maximum value very soon after noon, and decreases almost to zero shortly after sunset, and throughout the night.

By adding (9) and (10) we get, accurately (assuming that only photochemical changes occur, without convective transport),

$$\tfrac{1}{2}d(n_1+n_3)/dt = Q_2 - k_{11}n_1{}^2n - k_{13}n_1n_3. \qquad \ldots\ldots(16)$$

Here the terms on the right are very small compared with Q_3 and $k_{12}n_1n_2n$. These small terms govern the changes of n_1+n_3, which measures the number of *odd* atoms of oxygen, whether isolated or, as in the ozone molecule, combined with an atomic pair O_2. During the day there are isolated O atoms, and the number of O_3 molecules is reduced by an almost equal number; if this reduction, at noon when it is greatest, is by a fraction f of n_3, then $(n_1)_{noon} = fn_3$ very approximately, and if f were known, and Q_3, we could infer the value of $k_{12}n_2n$ or k_{12} from (15); in fact $k_{12} = Q_3/n_1n_2n = (Q_3)_{noon}/fn_3n_2n$. For example, taking $Q_3 = 4{\cdot}4\times10^9$ per c.c. (Wulf and Deming, case C, at 27·5 km.), $n_3 = 5\times10^{12}$, $n_2 = 10^{17}$, $n = 4\times10^{17}$ and $f = 1/10$, we get $k_{12} = 2{\cdot}2\times10^{-37}$; if $f<1/10$, k_{12} will be greater than this.

The terms on the right of (16) are zero at night, so that the changes in n_1+n_3 occur only by day, namely by dissociation of O_2 molecules, and by reunion of O atoms into pairs (O_2), according to either (6) or (7).

§ 7. THE SEASONAL OZONE VARIATION

If we write $y = n_1+n_3$, so that (16) becomes

$$\tfrac{1}{2}dy/dt = Q_2 - (k_{11}n - k_{13})n_1{}^2 - k_{13}n_1y,$$

and substitute for n_1 from (15), we get

$$\tfrac{1}{2}dy/dt = Q_2 - (k_{11}n - k_{13})Q_3{}^2/(k_{12}n_2n)^2 - k_{13}Q_3y/k_{12}n_2n. \qquad \ldots\ldots(17)$$

In this equation n, n_2 and the k's may be considered as constant, and Q_2 and Q_3 may be supposed to vary as $\cos(\pi t/t_0)$ throughout the hours of sunlight, from $t=0$ at dawn to $t=t_0$ at sunset; this cosine formula is, however, only a rough approximation (Chapman, 1931 a, figures 2–5).

Integrating the right-hand side of (17) throughout the hours of daylight on one day, we have, as the change of y during this time,*

$$(4t_0/\pi)Q_2' - t_0(k_{11}n - k_{13})(Q_3'/k_{12}n_2n)^2 - (4t_0/\pi)k_{13}yQ_3'/k_{12}n_2n, \qquad \ldots\ldots(18)$$

where Q_2' and Q_3' are the noon values of Q_2 and Q_3. If we identify y with the midnight value of n_3 (because $n_1=0$ at midnight), (18) represents the change in n_3 from one midnight to the next. This change will vary throughout the

* Chapman (1930 a) in his equation (17) omitted a factor 2 in the last two terms. His subsequent equations can be corrected (as is done here) by writing $2k_{11}$ and $2k_{12}$ in place of k_{11} and k_{12}. The present summary also gives revised values for his numerical estimates of the k's, consequent on the revised observed height of maximum ozone (§ 1) and on the use of the values Q_2 and Q_3 given by Wulf and Deming.

year, because Q_2' and Q_3' are greater in summer than in winter. From (18) it is not difficult to deduce the approximate formula

$$n_3 = a_0 + a_1 \cos\tau + b_1 \sin\tau \qquad \ldots\ldots(19)$$

for the (midnight) value of n_3 throughout the year; here τ denotes time reckoned from the vernal equinox, at the rate 2π per year. As has already been stated, no regular daily variation of n_3 has been observed, so that (19) very closely represents also the value of n_3 by day; it refers to moderate latitudes.

The value of a_0, the annual mean value of n_3, is found from (18) to be

$$\frac{k_{12}n_2nQ_2'}{k_{13}Q_3'} - \frac{\pi}{4}\frac{Q_3'}{k_{13}n_2n}\left(\frac{k_{11}n}{k_{13}} - 1\right). \qquad \ldots\ldots(20)$$

The first term here agrees with the equilibrium value (14) used by Wulf and Deming, if there we ignore the small term Q_2 in the sum Q_2+Q_3; in (20) the second term represents the departure from their equilibrium value, due partly to the daily variation of Q_2 and Q_3, and partly to the consideration of the reaction (5).

As regards the seasonal variation, (18) leads to the relation

$$b_1/a_1 = -Nk_{13}Q_3'/\pi^2 k_{12}n_2n = -Nk_{13}(n_1)_{\text{noon}}/\pi^2, \qquad \ldots\ldots(21)$$

where N denotes the number of seconds in a year; the negative sign of b_1/a_1 implies that the maximum ozone occurs *before* the vernal equinox. The observed seasonal variation corresponds to a small value of b_1/a_1; if the ozone maximum occurs a week before the equinox, b_1/a_1 will be about $-1/10$, but its value has not yet been accurately observed. If b_1/a_1 were known, and also Q_3' and (from the daily ozone variation) k_{12}, it would be possible to infer the value of k_{13}; for example, taking $b_1/a_1 = -1/10$, and also $f = 1/10$, and $Q_3' = 4\cdot4\times10^{19}$, as in §6, we find $k_{13} = 6\times10^{-20}$, so that $k_{12}/k_{13} = 4\times10^{-18}$. These values are merely illustrative, as f and b_1/a_1 are not known from observation; but the estimate of k_{12}/k_{13} does not differ much from one of the values inferred by Wulf and Deming in a very different way.

If b_1/a_1 is small, (18) leads to the formula

$$a_1 = (N/4\pi)k_{13}(n_1)^2_{\text{noon}}\alpha(k_{11}n/k_{13} - 1), \qquad \ldots\ldots(22)$$

where α measures the seasonal variation of Q_2' and Q_3', supposed to be proportional to $1+\alpha\sin\tau$; thus α is zero at the equator, and increases with the latitude. The sign of a_1 given by (22) corresponds to maximum ozone in spring, and minimum in autumn, as observed, if $k_{11}n > k_{13}$.

At Oxford α is approximately $\frac{1}{3}$, and a_1 is about 2×10^{12}; taking $(n_1)_{\text{noon}}$ as 5×10^{11} and n as 4×10^{17}, as in §6, (22) indicates that $k_{11}n$ much exceeds k_{13}, and that $k_{11} = 2\cdot3\times10^{-35}$. This estimate involves the *square* of the unknown n_1 or f (§6); if f were 1/100 instead of 1/10, k_{11} would become $2\cdot3\times10^{-33}$. If $k_{11}n/k_{13}$ is large, as is here suggested, (22) takes the much simplified form

$$a_1 = (N/4\pi)\alpha k_{11}n(n_1)^2_{\text{noon}}.$$

It may be hoped that laboratory researches will in time give the values of k_{12}, k_{11} and k_{13}, in which case, instead of inferring their values from the daily and seasonal variations of n_3, as here, these variations may be checked against the values calculated from the above photochemical theory. Any discrepancy

would presumably have to be attributed to convective transfer of ozone between different heights and between different latitudes.

Nevertheless it is satisfactory to note that these estimates of k_{12}, k_{11}, k_{13}, based purely on solar and meteorological data, all appear to be reasonable in order of magnitude. The conclusion that k_{11} exceeds k_{12} is consistent with the fact (pointed out to me by Dr. W. G. Penney) that O atoms are symmetrical, so that no condition as to orientation need be satisfied in the reaction (5), such as may be required in the case of (6). Dr. Penney also points out that when the third body M in these reactions is a nitrogen molecule (as will be the case, in the atmosphere, three times out of four), the conditions are favourable for absorption (into vibrational energy) of any energy surplus involved in the reaction.

§ 8. THE VARIATION OF THE OZONE DENSITY WITH HEIGHT

A theoretical discussion of the height-distribution of the ozone content has been given by Chapman (1930 b). He showed that at sufficiently great heights n_3 must be proportional to $n_1 n_2 n$; as all these three factors decrease upwards, the decrease of n_3 must there be extremely rapid. Moreover, n_1/n_2 must increase upwards, so that at great heights the oxygen must be more and more completely dissociated. The level at which (say) more than half the oxygen is present in the atomic form will depend chiefly on the value of k_{11}, and on the distribution of temperature at high levels.

The calculations of n_3 by Wulf and Deming ignored the reaction (5), which seems to be important in the main part of the ozone layer, and becomes all-important at higher levels. Even so, their results illustrate the very rapid decrease of n_3/n in the upper part of the ozone layer.

§ 9. THE LIGHT OF THE NIGHT SKY

Chapman (1931 c, 37) concluded, after examination of various possibilities, that the most likely source of the energy of the (non-auroral) light emitted by the night sky is the reaction (5).

REFERENCES

CHAPMAN, S., 1930 a. *Mem. R. Met. Soc.* **3**, 103–125.
CHAPMAN, S., 1930 b. *Phil. Mag.* (7), **10**, 345–352.
CHAPMAN, S., 1930 c. *Phil. Mag.* (7), **10**, 369–383.
CHAPMAN, S., 1931 a. *Proc. Phys. Soc.* **43**, 26–45.
CHAPMAN, S., 1931 b. *Proc. Phys. Soc.* **43**, 483–501.
CHAPMAN, S., 1931 c. *Proc. Roy. Soc.* A, **132**, 353–374.
CHAPMAN, S., 1937. *Phil. Mag.* (7), **23**, 657–664.
CHAPMAN, S. and MILNE, E. A., 1920. *Quart. J. R. Met. Soc.* **46**, 357–396.
GÖTZ, F. W. P., 1931. *Ergeb. Kosmischen Phys.* **1**, 180–235 (Leipzig: Akad. Verlagsges.).
GÖTZ, F. W. P., 1938. *Ergeb. Kosmischen Phys.* **3**, 253–325 (Leipzig: Akad. Verlagsges.).
MECKE, R., 1931 a. *Z. phys. Chem.*, Bodenstein-Festband, 391–404.
MECKE, R., 1931 b. *Trans. Faraday Soc.* **27**, 375–377.
MECKE, R., 1931 c. *Trans. Faraday Soc.* **27**, 359–375.
WULF, O. R., 1932. *Phys. Rev.* (2), **41**, 375.
WULF, O. R., 1934. *Phil. Mag.* **17**, 251–263.
WULF, O. R., 1935. *J. Opt. Soc. Amer.* **25**, 231.
WULF, O. R. and DEMING, L. S., 1936 a. *Terr. Mag.* **41**, 299–310.
WULF, O. R. and DEMING, L. S., 1936 b. *Terr. Mag.* **41**, 375–378.
WULF, O. R. and DEMING, L. S., 1937. *Terr. Mag.* **42**, 195–202.
WULF, O. R. and MELVIN, E. H., 1931. *Phys. Rev.* **38**, 333.

SOLAR PHYSICS

By A. HUNTER,
Royal Observatory, Greenwich

§1. INTRODUCTION

THIS article brings up to date a review by Dr. A. D. Thackeray (1940) in a previous volume of these *Reports*. Since that was written, astronomical work in the belligerent countries has been severely curtailed. In European countries most of the big telescopes used for astrophysical work have been dismantled for safety, a precaution which has been justified by the bombing of many of the great cities and by the destruction in ground fighting of no fewer than nine observatories in Russia alone. The solar physicist has been particularly fortunate: his work is done in friendly daylight, and often with small instruments which it has not been necessary to dismount for security. Add to this the fact that neutral countries, such as Switzerland, have long been in the forefront in solar research, and it becomes understandable that investigations of the Sun have figured largely in astronomical journals during the past three years.

§2. THE CORONAL SPECTRUM

The line spectrum of the inner corona, first observed three-quarters of a century ago, has presented an outstanding challenge to astrophysicists ever since. One by one the other unidentified celestial spectra—the D_3 line in the solar chromosphere, the N1 and N2 lines in galactic nebulae, the infra-red absorption bands in the major planets*—yielded to concerted attack, but the coronal spectrum remained a mystery until quite recently. The publication of a paper by Edlén (1942) now bids fair, in removing the mystery, to substitute for it a revolutionary conception of the physical conditions in the outer solar envelope. He attributes the lines to atoms of Fe, Ni and Ca in very highly ionized states corresponding to excitation temperatures of millions of degrees.

Early attempts to identify "coronium" by seeking wave-length coincidences with known spectra failed because they never explained more than a small minority of the observed facts, and it was virtually certain, when Bowen in 1935 identified "nebulium" as ionized oxygen and nitrogen, that coronium too would turn out to be some known element or mixture of elements excited in an unfamiliar way by a process which develops the coronal spectrum whilst suppressing the normal one. The most likely process of this type is, of course, ionization: astronomers in particular have good cause to remember that in successive stages of ionization an atom emits a series of quite distinct spectra. But ionization alone, it was thought, was not enough: if it were, surely one of the great variety of terrestrial modes of excitation known even fifteen years ago would have

* And now, also, the λ 4050 group in comets, produced in the laboratory and attributed to CH_3 by Herzberg (1942), and perhaps the λ 3556 band in the night sky, attributed to O_2 by Dufay (Swings, 1943).

revealed the origin of the lines? Hope was therefore centred on the metastable states of the commoner ions; forbidden inward transitions from these states were found to be involved in the production of the nebular lines, and by analogy it was thought likely that the coronal lines, too, were of this type. This need not hamper the search for their origin, since the energy levels concerned may be known quite well from permitted transitions, observed in the laboratory, which involve them. Nevertheless, when about five years ago a systematic search amongst the metastable levels of the ions of such atoms as Fe, Ca, Ti, A and many others had failed to give any coincidences with the coronal lines, this avenue of approach began to be viewed with disfavour. It was pointed out that although the low pressure in the corona would favour metastability, because the excited atoms would have little chance of unloading their energy by collisions, they should be easily *raised* by the all-pervading flood of solar radiation to normal electronic states, from which permitted transitions are possible to the ground level. It was further emphasized that the great width of the lines pointed to excited states of short lifetimes—quite the reverse of metastability.

But other lines of attack on the problem proved equally unsuccessful, and it was left to Edlén to propose a bold advance on the previous work, and to suggest that the coronal lines are indeed produced by forbidden transitions in ions of known elements, but that the stage of ionization is far higher than had previously been imagined. Abandoning the somewhat over-cautious approach of previous work, he examined the metastable levels of atoms which have lost from nine to fifteen electrons. Extreme excitation by high-voltage sparks between metal electrodes *in vacuo* can produce such ions (see, for example, Shenstone, 1938). Their spectra lie in the extreme ultra-violet close to the soft x-ray region, and are observed in grazing-incidence vacuum spectrographs; but, as excited in the laboratory, they contain of course only the permitted lines of the half-stripped ions. Nevertheless, term analyses of the spectra by Edlén and his co-workers at Uppsala have revealed that, in many of these ions, metastable levels do exist, and, further, that forbidden transitions from these levels will be expected to produce lines in the visible region of the spectrum. It seems that the first suggestion that the coronal lines might be of this type was made by Grotrian (1939), who observed that the wave-numbers of the coronal lines at 6374 and 7892 A. coincide with the doublet separations in the ground terms of Fe x and Fe xi. This suggested a search for analogous separations in other ions, a search for which the Swedish workers, with their unique knowledge of these higher spark spectra, were particularly fitted. At once Edlén discovered that the analogous term separations in the isoelectronic ions Ca xii and Ca xiii coincide with two other strong coronal lines (see table 1); and the conviction grew that he was searching on the right lines. At this stage, however, imperfections in the term analyses of the spectra of these highly-ionized atoms imposed limitations, and Edlén had to abandon established term differences for extrapolated ones. The method of extrapolation along isoelectronic sequences is, however, well understood and has been found reliable, so that once we accept his basic thesis that the whole of the coronal spectrum consists of forbidden transitions in highly-ionized atoms, any numerical coincidences which fall within the error of prediction may be regarded as probably significant.

The subsequent argument runs as follows. The strength of Fe x and Fe x suggests that iron is present in abundance, as in most other cosmic sources. The Fe vii lines which appeared in the spectrum of Nova Pictoris are, however, absent, but Ca xii and Ca xiii are strong; evidently we should seek amongst the higher Fe ions for further coincidences. The extrapolated term differences in Fe xii are not in the astronomically observable region, but the ground-doublet separation in Fe xiv points unmistakably to the strongest line visible in the corona, namely the green one at 5303 A., and the strong infra-red lines discovered some years ago by Lyot without an eclipse are attributable to Fe xiii. The next three ions have no deep metastable levels, and beyond that the excitation

Table 1

Wave-length (A.)	Intensity		Identification		Ionization potential (ev.)
	Grotrian	Lyot	Ion	Transition	
3328·1	1·0		Ca XII	$^2P_{3/2}$–$^2P_{1/2}$	589
3388·10	16		Fe XIII	$^3P_2$ –$^1D_2$	325
3454·13	2·3				
3600·97	2·1		Ni XVI	$^2P_{1/2}$–$^2P_{3/2}$	455
3642·87			Ni XIII	$^3P_1$ –$^1D_2$	350
3800·77					
3986·88	0·7		Fe XI	$^2P_1$ –$^1D_2$	261
4086·29	1·0		Ca XIII	$^3P_2$ –$^3P_1$	655
4231·4	2·6		Ni XII	$^2P_{3/2}$–$^2P_{1/2}$	318
4311·5					
4359					
4567	1·1				
5116·03	4·3	2·6	Ni XIII	$^3P_2$ –$^3P_1$	350
5302·86	100	120	Fe XIV	$^2P_{1/2}$–$^2P_{3/2}$	355
5694·42		1·5			
6374·51	8·1	28	Fe X	$^2P_{3/2}$–$^2P_{1/2}$	233
6701·83	5·4	3·3	Ni XV	$^3P_0$ –$^3P_1$	422
7059·62		4			
7891·94		29	Fe XI	$^2P_2$ –$^3P_1$	261
8024·21		1·3	Ni XV	$^3P_1$ –$^3P_2$	422
10746·80		240	Fe XIII	$^3P_0$ –$^3P_1$	325
10797·95		150	Fe XIII	$^3P_1$ –$^3P_2$	325

is no doubt too high. When the intensities are taken into account, the lines so far identified as due to iron atoms, from which 9, 10, 12 and 13 electrons have been stripped, make up nearly 90 per cent of the total coronal line emission. Extrapolation to the isoelectronic ions of nickel shows that the Ni analogues of all these Fe lines (except one outside the observable range of wave-lengths) appear in the coronal spectrum with intensities roughly one-tenth those of the Fe lines. This accounts for a further 7 per cent of the emission, so that Edlén's proposals attribute all but a small fraction of the observed line emission to forbidden transitions in highly-ionized calcium, iron and nickel. Many of the other elements known to be abundant in the Sun would not be expected to show in the corona, since at the appropriate ionization level their spectra either do not

contain forbidden lines of this type at all, owing to the lack of deep metastable levels, or, if they do, the lines fall outside the somewhat limited region accessible to observation. Atoms of K, Cr, Mn and Co, however, are well represented in the Sun, and their ions should show observable forbidden lines. Edlén attributes the fact that they are not observed to low abundance in the corona; and he points out that meteorites are rich in Fe and Ni in about the ratio suggested by the relative strengths of their coronal lines, but do not contain K, Cr, Mn or Co. The ions of other elements occurring in meteoritic matter, O, Mg, Al, Si and S, have no strong lines in the visible region. On this view the corona is not directly of solar origin at all, but is a cloud of meteoric gas; and the division between the chromosphere and the corona becomes a physically significant discontinuity.

There is much supporting evidence for Edlén's identifications. The only other source in which the coronal lines have ever appeared is RS Ophiuchi, a variable star which blazed up like a nova in 1933. It is known that novae in their later stages exhibit forbidden lines of highly-ionized iron. The atmospheric shells which they throw off must come in time to resemble the solar corona on a vastly increased scale, so that it is not surprising to find very faint forbidden lines of ionized iron emitted in the Sun's outer atmosphere. (They are, of course, very faint compared with the ordinary solar spectrum: it is only due to the coincidence that the apparent sizes of the Sun and Moon are almost equal that we ever became aware of the existence of a solar corona.) It is not to be expected that McLennan's feat of reproducing the auroral lines in the laboratory will be repeated with the coronal lines: as yet it is only possible to strip iron atoms to this extent long enough to enable them to emit permitted lines; to keep them in metastable states long enough for the emission of forbidden lines is beyond our present technique. That they can populate metastable levels for a time of the order of seconds in the corona is,[1] of course, due to the low pressure and consequent infrequency of collisions. The objection raised earlier that solar radiation will raise metastable ions to normal levels, whence permitted transitions can take place, is neatly disposed of in Edlén's theory by the fact that the ultimate absorption lines of these ions are in the soft x-ray region, where the radiation of the Sun is presumably very weak. It is also encouraging to find that previous classifications of the coronal lines on the basis of their physical behaviour and their intensity distribution in various coronal zones segregates them in a way which is now seen to be a function of excitation level. In particular, the great difference between the behaviour of the red (6374 A.) and of the green (5303 A.) lines, which frequently reach maximum intensity at quite different parts of the corona, is accounted for by the fact that the emitting ions are Fe X and Fe XIV respectively. Finally, the theory not only accounts for the emission lines observed in the inner corona, but also throws light on the origin of the continuous spectrum of its outer regions. This may well be due to scattering of sunlight by the electrons stripped from the outer shells of the Ca, Fe and Ni atoms responsible for the emission lines. If we enquire how the stripped atoms and the electrons can co-exist, we are, of course, asking the fundamental question which underlies the whole hypothesis: what is the source of this extremely high excitation?

This is indeed the main difficulty the theory encounters. Many solar and extra-solar effects are now recognized to require much more energy than can be furnished by a black body of the Sun's size at a temperature of 6000° K. (see e.g. pp. 6–8 of this volume), but no other phenomenon requires an excitation temperature of the order of two or three million degrees. It is interesting, however, that the great width of the coronal lines, hitherto regarded as evidence that they were not forbidden transitions with long-lived upper states, is not so surprising if the "temperature" in coronal regions reaches this order of magnitude. It may be, on the other hand, that turbulent velocities known to exist in the corona are enough to account for the diffuse character of the lines: Waldmeier (1940a) at Arosa has observed directly, with a coronagraph, velocities of 10 km./sec. (0·2 A. at 6000 A.). This is the more likely since there seems some evidence from interferometer observations (Fessenkov, 1942) that the widths are not so great as has hitherto been believed. A more disturbing difficulty is encountered in Waldmeier's observation (1939) that the green line at 5303 A. sometimes shows self-absorption. If this is indeed so, the proposed identification with forbidden lines can hardly be upheld: only in intense electrical fields can such lines be absorbed, and presumably no such fields can exist in low-pressure atmospheres like the corona. But the observation is, strictly speaking, of an apparent doubling of the line; this may be a true doubling, and thus due to some other cause than self-reversal.

As regards the high excitation, the position in which we now find ourselves, with the outer envelope of the Sun more highly excited than its inner atmosphere, recalls the state of affairs existing just after the last war, before the ionization theory gave low pressure its proper place as a factor in excitation. Another Saha may explain the facts with equally satisfactory results. On the general point of the theory's plausibility, it must be admitted that only something of an unusual nature could have much chance of explaining the facts of coronal line emission; all the more conventional avenues of approach have been explored over and over again during the past twenty years. The question can be settled by thorough observation of the extreme ultra-violet spectra of Fe XIII and Fe XIV and the isoelectronic nickel ions. A definitive term analysis based on the permitted ultra-violet lines should not be impossible with present technique. If the theory is sound, the number of lines directly identified can then be increased from the present four to at least eleven, and all doubts regarding the validity of Edlén's extrapolations will be set at rest.

Meanwhile, astronomical observations can be continued, at eclipses, with coronagraphs such as those at the Pic du Midi and at Arosa, and perhaps also, in the future, with the coronaviser (Babcock, 1942). It is becoming increasingly clear that there are large day-to-day variations in the relative intensities of the coronal lines. Observations made during a total eclipse are therefore not necessarily entirely representative, and lines which appear at one eclipse may be absent at another. More careful scrutiny of old eclipse plates would therefore probably be well worth while, and some lines previously rejected, on what are now seen to be insufficient grounds, as not of coronal origin can be added to the list of observed lines. A recent paper (Petrie and Menzel, 1942) lists at least twelve new lines observed at the 1936 eclipse. Any of these might yield useful

clues in the search for identifications. If a combined attack on the problem from laboratory and observatory should confirm Edlén's proposals, it is not too much to say that a new era will have opened in solar physics, and, indeed, in astrophysics as a whole.

§3. BRIGHT CHROMOSPHERIC ERUPTIONS AND THEIR GEOPHYSICAL EFFECTS

The study of bright chromospheric eruptions continues apace. That they are responsible for short-wave radio fading by causing abnormal ionization below the *E* layer of the ionosphere is now established beyond doubt, though the mechanism by which they cause this ionization is by no means understood. Study of the visual spectra of the eruptions gives no conclusive answer to the problem. It suggests that line emission due to hydrogen at 1215 A. and to helium at 584 A. must dominate the ultra-violet spectrum. Unfortunately, radiation at 1215 A., though it can probably penetrate below the *E* layer, cannot directly cause ionization in any of the atmospheric gases there, since its wavelength is too great; whilst the 584 A. radiation, though it can ionize all constituents of the atmosphere, for that very reason cannot penetrate to the required level. In a review of other possibilities, Preston (1940) points out that higher members of the Lyman series, particularly numbers 4, 5 and 7 at 949·7, 937·8 and 926·2 A. respectively, might be responsible for the ionization. But the phenomena of radio fade-outs strongly suggest that the ionizing radiation is qualitatively different from ordinary solar radiation, which is certainly intense in the Lyman lines. The fact that the normal *E* and *F* layers are hardly disturbed by solar eruptions makes it probable that we must look for some radiation which does not normally reach us from the Sun in any appreciable quantity. Moderately hard X rays may be involved, but independent evidence of their production in eruptive regions of the solar chromosphere is lacking.

Since nearly all bright eruptions occur in the neighbourhood of sunspots, it is difficult to tell whether the long-known correlation between magnetic storms and the passage of individual spots across the Sun's central meridian is an effect of the spot itself or of an associated bright eruption. The data which have accumulated during the last maximum of solar activity and the subsequent years strongly support the view that at any rate with the "great" storms the effective phenomenon is a chromospheric eruption occurring within 45° of the centre of the disk about twenty-four hours before the beginning of the storm. The time-lag and the marked dependence on position on the disk sharply differentiate this mechanism from that producing the radio fade-outs discussed above. We are confronted here with the approximately radial emission from the eruption of particles with speeds up to 1600 km./sec. It is significant that this is the limiting speed set by Milne (1926) for the ejection of atoms by his "vertical instability" process, in which an atom, originally in equilibrium between radiation pressure and gravity, experiences, when once it starts to move outwards, more and more intense radiation as the Doppler effect of its outward velocity causes it to absorb further and further on the violet wing of its absorption line. The resulting acceleration enables it to escape the gravitational field before diminishing owing to inverse-square weakening of the radiation. It

may well be that the trigger initiating the unstable outward movement is a chromospheric eruption.

A more recent theory of Alfvén (1940) attributes magnetic storms and aurorae to a neutral stream of electrons and ions repelled diamagnetically from the Sun. The electrons stream past the Earth without being able to penetrate an unsymmetrical "forbidden" zone of mean diameter 10^{10} cm. surrounding it. Ions can, however, enter this region; they produce a positive space charge, on the side of the forbidden zone nearest the Sun, which repels further entering ions along the lines of force to the daylight side of the Earth, the atmosphere of which they reach in the auroral zones. In the shadow of the forbidden region the loss of ions results in a negative space charge repelling electrons along the lines of force to the night side of the Earth. The resulting magnetic disturbances are shown to bear remarkable resemblances to the actual state of affairs if suitable values are given to the energy of the particles leaving the Sun and to their rate of emission. In a criticism of this mechanism, Cowling (1942) shows that the accumulation of ions on the boundaries of the forbidden zone considerably modifies its size, and that electrostatic forces prevent more than a slight separation of electrons from ions. It seems possible, however, that in a sufficiently dense cloud the leading particles will shield the body of the stream from diamagnetic repulsion, and the particles can penetrate the atmosphere.

It seems clear that none of the present theories of the causation of magnetic storms by solar corpuscles is entirely satisfactory, and at the present stage we may state with fair confidence only that many chromospheric eruptions are the simultaneous source of radiation travelling in all directions at the speed of light, and of particles emitted roughly radially at effective speeds up to 1600 km./sec. The radiation consists, in the visible region, of line emission chiefly in the lines of H, He and certain metals, and, in the ultra-violet, of emission of unknown quality capable of causing abnormal ionization below the *E* layer on the sunlit side of the Earth. This increased conductivity in the *D* region of the ionosphere causes abnormal attenuation in short radio waves and also, via the Balfour Stewart atmospheric-dynamo mechanism, small characteristic disturbances in the magnetic elements in the same sense as the normal diurnal variation. These effects last only as long as the visible flare, i.e. an hour or less. The corpuscular emission causes widespread magnetic storms and aurorae if it reaches the Earth; this only happens in general if the generating flare occurs near the centre of the Sun's disk as seen from the Earth.

It seems possible that the intensity of the eruption in Hα is a measure of its effectiveness in producing geophysical consequences. Until recently, visual estimation on a rough numerical scale was used to distinguish between very bright (3), bright (2) and less bright (1) eruptions. An interesting method of measuring the brightness of an eruption has been developed by Waldmeier (1940b). In a spectrohelioscope, the bright marking is visible in an emission line, say Hα, against the background of undisturbed disk, not only at the centre of the line but also throughout definite wave-length ranges on either side. The total visibility width thus defined corresponds well with intensity observations and has the advantage that it is independent of the conditions of "seeing". The normal Hα flocculus from which the typical eruption develops has a visibility

width of just under 1 A. Where the eruption starts, a sudden broadening from 1 A. to 3 A. occurs in about five minutes, followed by a slower decline to normal. The actual central intensities in Hα can be measured at the limb by opening the viewing slit of the spectrohelioscope until the limb appears as bright as the eruption. By integrating, over this slit width, the contour of the Hα emission and that of the surrounding disk, and equating the two results, the true central intensities are found to reach 0·6 for eruptions of intensity 1, 1·5 for intensity 2, and 2·3 for intensity 3, the unit being the intensity of the continuous spectrum close to Hα. On correlating these central intensities with the geophysical results it may be stated in general terms that geophysical effects are frequently not marked enough to be noted if the intensity of emission in Hα does not rise above that of the neighbouring continuum, but that if the absorption line is much more than filled up, radio and geomagnetic effects are to be expected.

As regards the magnetic observations, it must not be forgotten that many magnetic storms, especially those observed near solar minimum and those which recur in the solar rotation period of twenty-seven days, cannot be connected with any photospheric or chromospheric activity. They are certainly not due to either sunspots or chromospheric eruptions. A recent series of observations by Waldmeier (1942) is of great importance in tracking down the origin of these disturbances. In mapping the form and intensity of the corona by means of day-to-day coronagraph observations, he has found that "C-regions" showing abnormal intensity in λ5303 occasionally appear at the Sun's limbs, and that about seven days later (or earlier, if the observation is made at the west limb), marked geomagnetic disturbances are likely to set in. The lag corresponds, of course, to a quarter of a solar rotation, so these active coronal regions seem to cause magnetic activity when they cross the central meridian in much the same way as chromospheric eruptions do. Out of the ten magnetic storms in 1939 and 1940 (years of declining solar activity) before or after which coronal observations were possible, seven were found to be associated with C-regions. Some of the remainder were also no doubt so associated, but the coronal activity either began after the region left the east limb or ceased before it reached the west limb. In their general characteristics these C-regions bear a remarkable resemblance to the M-regions whose existence on the Sun was postulated by Bartels some years ago to account for periodic magnetic storms by a process of long-continued corpuscular emission, and further observations will be awaited with interest.

A contribution to the vexed question regarding the connection between cosmic-ray intensity and magnetic storms comes from Dupérier (1942), who was making hourly observations of cosmic-ray intensity throughout the magnetic storm of March 1, 1942. This storm was preceded by a brilliant and prolonged chromospheric eruption observed on February 28 between 12.42 and 15.25 Universal Time in a large sunspot group due to pass the central meridian on February 28.8 (Newton, 1942a). Spectrohelioscopic observations were intermittent owing to cloud, but a short-wave radio fade-out started at 12.00 and continued on some channels till 20.00. "Crochets" of characteristic appearance but of abnormal range occurred on Abinger magnetograms at 12.01. Nineteen and a half hours later, a great magnetic storm began abruptly, lasting for 20 hours and showing a maximum range of 512×10^{-5} gauss in H, 176×10^{-5} gauss in V,

and 86′ in *D*. Dupérier's apparatus, which consists of nine Geiger counters arranged to register three-fold coincidences, and counts mesotrons at the rate of 400 per minute by means of a scale-of-two counter with three pairs of thyratrons, showed a slight fall of intensity shortly before the beginning of the storm, and during its progress an unprecedented drop of 12 per cent was recorded. The connection between the two types of disturbance is very imperfectly understood; bigger storms have occurred with much less violent disturbance of cosmic-ray intensity, one in particular (that of 1937 August 21–25) with practically no effect on cosmic rays. It is evident that much more observational data must be available before we can deduce a comprehensive theory relating changes in the Sun's photosphere, chromosphere and corona to the corresponding changes in the Earth's upper atmosphere.

§4. MOTION OF SOLAR PROMINENCES

Recent progress in this field has almost all been observational; we are still without a theoretical explanation of the motions observed, and even over the interpretation of some of the observations there is still controversy. In an investigation of the ascending motion of nineteen knots or features in an eruptive prominence, Waldmeier and Beck (1942) use coronagraph photographs instead of the conventional spectroheliograms. This method has the advantage that fine detail can be recorded without any fear of losing the most interesting part of the picture, as happens where large line-of-sight velocities shift the image off the selecting slit of the spectroheliograph. The observations confirm those of Pettit and his collaborators that sudden changes of velocity do occur in the observed motions, distance-time diagrams being best represented as a series of discontinuous straight lines. The projected trajectories of the knots are themselves shown as short straight lines, many of them parallel, at the junctions of which sudden changes of direction have occurred, alternately to right and to left of the general direction of ascent. It is particularly puzzling that the direction and velocity changes coincide only rarely in space or time. In contradiction of Pettit's second law of prominence motion, the velocities change by non-integral factors varying from 0·46 to 2·63.

A different method of attack has been used in this country by Newton (1942 b) and Ellison (1942 a) independently. These observers use spectrohelioscopes and measure line-of-sight velocities of prominences (i.e. "dark markings") directly instead of deducing velocities from distance-time diagrams. The advantage of this method is that whilst heights travelled can still be derived by integration, accelerations, and thence the forces acting, can be found by a single differentiation instead of the double process needed in reducing the limb observations. It has, of course, the disadvantage that only one component of the motion is observed, though with observations of prominences near the centre of the disk this component is perhaps the most interesting one, lying as it does in the direction of solar gravity.

Newton's observations are made on the dark markings associated with chromospheric eruptions; Ellison's on these and on the dark flocculi observed near spot groups at all times. The markings are probably allied to, if not identical with, the "surges" seen at the limb on motion pictures of prominences. About

one-half of those seen near eruptions have life histories closely resembling one another. Whilst the eruption is mounting in intensity, the dark marking close by shows a rapidly increasing outward velocity which rises to a maximum of about 150 km./sec. near the peak of intensity of the bright eruption. A rapid decline of velocity follows, merging into a constant retardation roughly one-quarter of that due to solar gravity. This stage lasts for nearly ten minutes as the absorbing gas rises to the top of its trajectory and the associated bright eruption fades. Before zero velocity is observed, however, the marking fades from view at about -30 km./sec., and another takes its place, showing an inward velocity of about $+30$ km./sec. There is often an appreciable interval during which the same point on the disk exhibits both inward and outward velocities simultaneously, i.e. the absorption is on both sides of Hα at once. The inward velocity increases with about the same acceleration, until the marking finally fades at a velocity of about $+80$ km./sec. At no stage, ascending or descending, are there observed the sudden changes of velocity of Pettit's first law of prominence motions, although the observations are quite precise and frequent enough to detect them if they exist.

The physical picture underlying these velocity observations is difficult to interpret, though an attempt by Hunter (1942) explains the most puzzling features, namely, the diminished acceleration and the simultaneous inward and outward velocities, by assuming that a stream of matter ejected continuously from the eruption for about seven minutes falls back under the action of gravity alone. Thackeray (1942) points out that caution is necessary in interpreting line-shifter readings on the spectroheliscope as Doppler velocities, and Ellison (1942b) emphasizes the rôle played by contrast in these observations. The immediate need is for intensity-calibrated high-dispersion spectrograms, taken throughout the history of a bright eruption, on which the contours of the absorption lines can be measured. The idea that the observed shifts may be due to changes of contrast or of contour can then be dismissed or accepted as the observations demand.

It is perhaps worth remarking here that spectroheliscopic observations on the bright matter of an eruption show that no marked outward velocity exists. The name "chromospheric eruption" is thus somewhat of a misnomer, in so far as the words suggest the outpouring of material in volcanic fashion, and the tendency is growing to refer to the phenomenon more non-committally as a "solar flare" (Hunter, 1943). In a recent paper Ellison (1943) shows that small fleeting redward shifts (up to 0·3 A.), which are positively correlated with intensity of eruption, may exist; but he concludes that these are the results rather of a slight absorption in the violet wing of Hα than of a true Doppler shift of the whole contour. Here again there is an uncomfortable suspicion that the spectroheliscope is, in fact, incapable of showing big radial velocities in some bright solar markings; when they are shifted by Doppler effect they are compared with the bright wings of the line in the disk background, and lose their contrast. Contour observations are again badly needed to settle the point.

Whilst dealing with this subject of prominences in the neighbourhood of solar flares, it should be mentioned that there is no direct observational evidence for the existence of chromospheric matter ejected at the 1600 km./sec. demanded

both by observations of the time-lag of magnetic storms and by Milne's theory of vertical instability in the chromosphere (see § 3 above). The velocity of escape at the surface is 620 km./sec.; this has occasionally been exceeded in observed prominences, but the surges which are believed to be associated with chromospheric eruptions rarely give maximum velocities exceeding 200 km./sec.

§ 5. CONVECTION IN THE SUN

In the previous *Report* on this subject (Thackeray, 1940), emphasis was laid on the importance attached in recent years to the effect of convection in the solar atmosphere. In a recent paper Eddington (1941) examines the layer near the surface where hydrogen is in the critical stage of ionization. This layer, which on the equilibrium theory extends from the 9000° to the 15,000° K. levels, is convectively unstable because the specific heat is large and the adiabatic constant much below the usual value 5/3. Thus the equilibrium theory breaks down; ionization equilibrium is reached slowly, and for quite small convection speeds the state of ionization fails to keep pace with the changing temperature and pressure of the material. In descending currents the ionizing photons entering the lower layer cannot ionize the material as fast as it is descending, whilst in ascending currents the net rate of recombination cannot free photons rapidly enough to prevent the arrival at the upper layer of hydrogen which is still mostly ionized. This occurs at current speeds too low to affect the dynamical and thermodynamical equations.

This breakdown of ionization equilibrium has important consequences. First of all, the original zone of convection is considerably extended. Secondly, the ionization gradient is flattened, though not, of course, enough to destroy the motive power of the currents; the actual convective speed will increase until the gradient is just sufficient to maintain convective instability. Last, and perhaps from the observational viewpoint most important, is the effect on the opacity of the convective layers. For radiation of energy below the ionization potential of hydrogen, 13·5 ev. ($\lambda > 911$ A.), the absorption by both ascending and descending currents will be small, since they mainly consist respectively of ionized and ground-state hydrogen; and such absorption is, of course, from excited levels of the neutral atom, which are eliminated by convection at even slower speeds than are needed for the main process. Radiation beyond the ionization limit will be strongly absorbed by the descending currents, but the ascending currents will be transparent to it. Taking both columns together, the convection will cause a substantial increase in the transmission by segregating, horizontally, the opacity to high- and low-frequency radiation, in much the same way as two windows, one of red and the other of green glass, transmit more light when side by side than if they consisted of half-thicknesses of red and green glass superposed. As a result, an abnormal quantity of high-frequency radiation from the inner regions will penetrate to layers just below the photosphere. In addition to this, similar radiation will be generated there by the recombination of hydrogen ions with electrons at the top of the convective columns. Thus we may expect a considerable excess of ultra-violet radiation over the black-body value. Another *Report* in this volume (see p. 5) gives observational evidence for the existence of just such an excess.

Eddington's discussion, as he pointed out at the time, neglects the effect of collision ionization. A recent paper by Schwarzschild (1942), using a new calculation of the effective cross-section for an ionizing collision between an electron and a hydrogen atom, makes it clear that collision ionization in the lower part of the convective zone will supplement photo-ionization to a considerable extent. At first sight this would appear to invalidate Eddington's results, but a further paper (Eddington, 1942) shows that, nevertheless, for other reasons a large difference in the state of ionization will exist between upward and downward streams. The previous conclusion that recombination at the upper limit of the convective zone will release an excess of high-frequency radiation in the outer layers of the Sun thus remains qualitatively unchanged, and we are again led to the conclusion demanded by observation, namely, that the ultra-violet spectrum of the Sun is much more intense than that of a black body at 6000° K.

REFERENCES

Alfvén, H., 1940. *Stockholm Vet.-Ak. Handl.* **18**, no. 9.
Babcock, H. W., 1942. *Astrophys. J.* **96**, 242.
Cowling, T. G., 1942. *Terr. Magn. Atmos. Elect.* **47**, 209.
Dupérier, A., 1942. *Observatory*, **64**, 190.
Eddington, A. S., 1941. *Mon. Not. R. Astr. Soc.* **101**, 177.
Eddington, A. S., 1942. *Mon. Not. R. Astr. Soc.* **102**, 154.
Edlén, B., 1942. *Ark. Mat. Astr. Fys.* **28** B, 1.
Ellison, M. A., 1942 a. *Mon. Not. R. Astr. Soc.* **102**, 11.
Ellison, M. A., 1942 b. *Observatory*, **64**, 300.
Ellison, M. A., 1943. *Mon. Not. R. Astr. Soc.* **103**, 3.
Fessenkov, B. G., 1942. *J. Phys. U.S.S.R.* **6**, 1.
Grotrian, W., 1939. *Naturwissenschaften*, **27**, 214.
Herzberg, G., 1942. *Astrophys. J.* **96**, 314.
Hunter, A., 1942. *Observatory*, **64**, 201.
Hunter, A., 1943. *Observatory*, **65**, 14.
Milne, E. A., 1926. *Mon. Not. R. Astr. Soc.* **86**, 459.
Newton, H. W., 1942 a. *Observatory*, **64**, 260.
Newton, H. W., 1942 b. *Mon. Not. R. Astr. Soc.* **102**, 2.
Petrie, W. and Menzel, D. H., 1942. *Astrophys. J.* **96**, 395.
Preston, W. M., 1940. *Phys. Rev.* **57**, 887.
Schwarzschild, M., 1942. *Mon. Not. R. Astr. Soc.* **102**, 152.
Shenstone, A. G., 1938. *Rep. Prog. Phys.* **5**, 221.
Swings, P., 1943. *Astrophys. J.* **97**, 72.
Thackeray, A. D., 1940. *Rep. Prog. Phys.* **7**, 160.
Thackeray, A. D., 1942. *Observatory*, **64**, 204.
Waldmeier, M., 1939. *Z. Astrophys.* **18**, 241.
Waldmeier, M., 1940 a. *Z. Astrophys.* **20**, 195.
Waldmeier, M., 1940 b. *Z. Astrophys.* **20**, 46.
Waldmeier, M., 1942. *Z. Astrophys.* **21**, 275.
Waldmeier, M. and Beck, B., 1942. *Z. Astrophys.* **21**, 286.

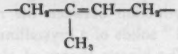
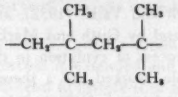
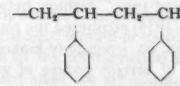
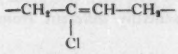
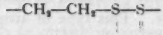
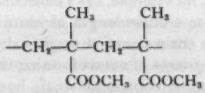
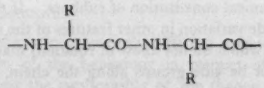
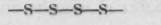
THE STRUCTURE AND ELASTICITY OF RUBBER

By L. R. G. TRELOAR,
The British Rubber Producers' Research Association

§1. INTRODUCTION

In its ability to undergo a very large reversible deformation natural rubber must be considered as a member of a large and rapidly increasing family of substances of widely diverse chemical constitution, a few of which are shown in table 1. Not all of these show rubber-like elasticity under normal conditions, polystyrene and polymethyl methacrylate, for example, become elastic only on heating, and are hard glass-like solids at ordinary temperatures; gelatin is elastic only when swollen with water. For all rubbers the range of

Table 1. Some rubber-like materials

Natural rubber	
Polyisobutylene	
Polystyrene	
Polychloroprene (" Neoprene ")	
Polyethylene disulphide (" Thiokol A ")	
Polymethyl methacrylate	
Gelatin	
Elastic sulphur	

temperature over which rubber-like behaviour is observed is limited. The lower limit is associated with the transformation to the glassy or crystalline solid state; the upper limit is associated with the transition to the viscous liquid state (unless chemical breakdown intervenes). In many cases, such as natural rubber, the upper limit may be raised by vulcanization, whilst in others, e.g. methacrylate, an extension of the range to lower temperatures may be achieved by the incorporation of a "plasticizer".

In this *Report* an attempt will be made to show how both the elastic and the other states of a rubber are related to the properties of the molecule and the arrangement of the molecules in the material.

§2. ELASTIC MOLECULES

Most of the theories put forward to explain the elasticity of rubber have related it directly to the structure of the molecule, the molecule thus being regarded as the fundamental elastic unit. But whereas many of the earlier theories attributed the elasticity to attractive forces between certain atoms or groups of atoms along the molecular chain, more recent theories explain the phenomenon in terms of the thermal motion of the chain atoms. This change in view-point, which represents nothing less than a revolution in thought, revealing a fundamental distinction between the rubber-like type of elasticity and the elasticity of "hard" solids of a crystalline or glassy character, took place about ten years ago. The kinetic theory of elasticity was first clearly expressed by Meyer, von Susich and Valko (1932), and was developed mathematically by Kuhn (1934, 36) and by Guth and Mark (1934). At the same time a similar concept was in process of evolution in the minds of other workers. Griffith (1934), for example, worked out a theory according to which the molecules rotated in "skipping-rope" fashion between their points of interconnection, and though Mack (1934) attributed the elasticity to the attractive forces between hydrogen atoms in an irregularly-kinked molecule, he was aware that the postulated ability of certain groups of atoms to rotate about the single C—C bonds as axes would lead to an enormous number of molecular configurations—a concept not essentially different from that underlying Kuhn's statistical treatment.

When we look at the molecular formulae listed in table 1, we find in all cases a chain-like structure. This type of structure is characteristic of all rubbers. The molecular weight of these long-chain molecules is invariably very high. Natural rubber, for example, has a molecular weight of about 300,000 (Gee, 1940), corresponding to a chain length of some 20,000 carbon atoms. Moreover, in all rubbers the chain contains single valence bonds, permitting relatively free rotation of the groups of atoms forming the chain. This characteristic—a long chain-like structure, containing single bonds—is the only common property of the chemical constitution of rubbers. If this characteristic is present there may be a wide variation in other features of the molecule. Thus, for example, double bonds (which do not allow of rotation) may or may not be present, there may or may not be side groups along the chain, the chain may be formed of carbon atoms, or it may be formed of other atoms (e.g. sulphur), or of an alternation of different atoms.

The kinetic theory of Kuhn (1934) idealizes the molecule, stripping it, so to speak, of all chemical properties, and treating it simply as a chain of freely-rotating links. This will be made clear from figure 1, representing a paraffin chain. If C_1, C_2 are the first two carbon atoms, the third atom C_3 lies on the rim of a cone formed by the rotation of C_2C_3 about C_1C_2 as axis, where $C_1C_2C_3$ is the valence angle ($109\frac{1}{2}°$). The fourth atom, C_4, lies similarly on the rim of a cone whose vertex moves round the first cone, and so on. We thus arrive at a randomly-fluctuating, irregularly-kinked form of the molecule, such that

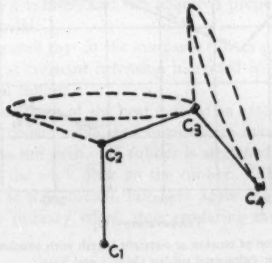


Figure 1. Rotations about bonds in paraffin molecule.

the average distance between the ends is only a small fraction of its outstretched length. It follows that if such a molecule is extended and then released, the thermal motion of the chain atoms will tend to restore it to its statistically most probable length. The elasticity is, therefore, a function of the statistical form of the free molecule, and arises from the heat motion of its component atoms.

Some of the recent developments in the mathematical theory of elasticity will be considered later; at the present stage we are concerned only with the general picture of the elastic rubber molecule which this theory enables us to draw.

§3. EXPERIMENTAL EVIDENCE FOR THE KINETIC THEORY

The kinetic approach leads directly to certain important results, which have been discussed in an earlier *Report* (Pelzer, 1939). These are:

(i) The reversible extension of rubber, at constant temperature, involves no change in its internal energy. The work done on the rubber is transformed into heat, and there is a resultant reduction in entropy on extension.

(ii) In rubber maintained at constant extension, the elastic tension is proportional to the absolute temperature.

In both these respects rubbers behave in an entirely different manner from ordinary elastic solids. Investigation of the behaviour in respect of either (i) or (ii) may therefore be used to test the applicability of the theory. In a careful study of the stress-temperature relations in vulcanized rubber, Meyer

and Ferri (1935) found that, provided that irreversible crystallization and plastic flow effects were avoided, the prediction of the theory was fulfilled. Figure 2, taken from their work, shows a proportionality between tension and temperature over a very wide range of temperatures. Other workers, also, have found the

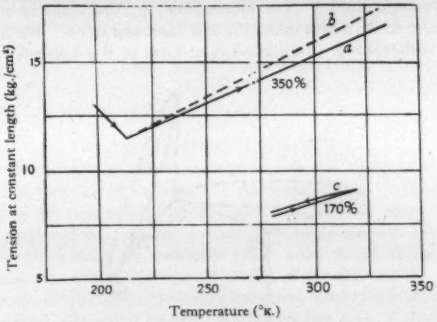


Figure 2. Variation of tension at constant length with absolute temperature for vulcanized rubber (Meyer and Ferri).
(*a*) 350 % elongation, experimental curve.
(*b*) 350 % elongation, corrected for thermal expansion.
(*c*) 170 % elongation.
(*Reproduced, by kind permission, from* ' *Helvetica Chimica Acta* ', **18**, 570.)

tension to increase with increasing temperature, but the relationship was not always a direct proportionality; the departures from this simple relationship are indicative of changes in internal energy, superimposed on the simple entropy effect. The most complete experimental investigation of such changes is

Table 2. Internal energy and entropy changes on stretching.
(From Wiegand and Snyder's curves.)

Extension per cent	F (gm. wt.)	$(\partial U/\partial l)_T$ (gm. wt.)	$T(\partial S/\partial l)_T$ (gm. wt.)
158	70	− 20	− 50
288	104	+ 24	− 80
376	121	+ 5	−116
462	113	−142	−255
548	131	−240	−371
632	156	−170	−326
718	250	+390	+140

probably that carried out by Wiegand and Snyder (1934), who covered the whole range of extension of vulcanized rubber. Table 2 gives the relative internal energy and entropy changes $(\partial U/\partial l)_T$ and $T(\partial S/\partial l)_T$ per unit extension derived (Treloar, 1942) from their data.

Wiegand and Snyder recognize three distinct regions in the extension curve. In the first region (0-350% extension) there is a reduction in entropy, accompanied by a relatively small increase in internal energy, in approximate agreement with the simple theory; in the second region (400–700% extension) there is a large reduction in internal energy, whilst in the third region (>700%) the energy and entropy terms both become positive. It is clear that in the second region a spurious effect is introduced by crystallization (see §5), the internal energy changes resulting from this cause being sufficient to obscure the changes relating directly to the elastic extension. In the third region the rubber has ceased to behave like a rubber, and has acquired properties comparable with those of a crystalline solid.

Meyer (1936) has stated that in the inorganic rubber phosphonitrilic chloride $(PNCl_2)_n$ the tension at constant extension increased with temperature, in the same way as in natural rubber.

In the direct measurement of the heat evolved on extension the experimental difficulties have been considerable, and quantitative data which might be used to check the theory do not exist. If rubber is stretched isothermally, the heat evolved should equal the work done on the rubber. This reversible evolution of heat on extension is well known, but here again the heat of crystallization tends to obscure the primary effect, thus rendering the interpretation of the data difficult.

§4. THE MOLECULAR STRUCTURE OF RUBBERS

So far the elasticity of the molecule has been considered without relation to the arrangement of molecules, and the forces between molecules, in the actual material. In passing on to these other considerations we see at once that certain other conditions have to be satisfied, if the bulk material is to be elastic. We are thus led to the following statement of the essential conditions for rubber-like elasticity, based primarily on the work of Busse (1932) and Treloar (1940):

(i) The presence of long-chain molecules, possessing freely-rotating links.

(ii) Weak secondary forces between the molecules.

(iii) An interlocking of the molecules at a few places along their length to form a 3-dimensional "network".

The first of these conditions has already been discussed. The second is necessary if the molecules are to have that freedom of movement which the kinetic theory of elasticity requires. The third arises from the general consideration that if the molecules are to be under tension they must be connected more or less firmly with each other, so that a continuous structure is formed, but these interconnections must not be so frequent that the freedom of motion of the chain atoms is impeded.

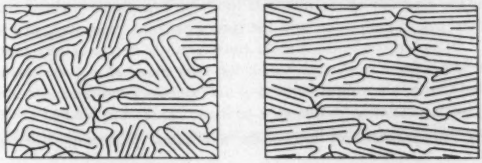
The ideal rubber therefore consists of an irregular network of long elastic molecules, connected together by chemical bonds. The portions of molecules between points of linkage are completely free to move among their neighbours, in the same way as the molecules of a liquid, subject only to the restrictions imposed by their forming part of a chain. Well-vulcanized natural rubber

approximates to this ideal at ordinary temperatures. In this case the cross-links are generally assumed to be chemical bonds resulting from the combination of the rubber with sulphur. In unvulcanized rubber, on the other hand, the interconnections are not thus sharply defined, but are probably cohesions of varying strength formed as a result of complex local entanglements between the hydrocarbon chains (Treloar, 1940). Unvulcanized rubber, therefore, shows complex time effects, due to the breaking down and re-formation of entanglements under strain, and, at higher temperatures, a plastic or irreversible deformation, effects which are much less noticeable in vulcanized rubber.

§ 5. CRYSTALLIZATION

Reference has already been made to crystallization in connection with the phenomenon of elasticity. Much work has been devoted to the elucidation of the mechanism of the process of crystallization, and information has been obtained which throws a good deal of light on the molecular states of rubber.

It has been known for many years that unvulcanized rubber becomes hard and inextensible if kept at 0° c. or a lower temperature for a few days. It is known also that raw rubber after stretching to its fullest extent will, under suitable conditions, remain extended for an indefinite time, but that retraction to the original length will occur if it is subsequently warmed to a certain sharply-defined temperature. These effects have been shown by x rays to be due to crystallization. Some, but not all, synthetic rubbers crystallize on stretching.



(*a*) Unstretched. (*b*) Stretched.

Figure 3. Molecular structure of crystalline rubber (diagrammatic). The parallel bundles represent crystallites.

At 90° c. vulcanized rubber shows crystallization only when stretched beyond 500 % extension (Field, 1941*a*). Hence crystallization, though important, must be regarded as a secondary effect, connected only indirectly with the phenomenon of elasticity.

The molecular structure of (*a*) unstretched and (*b*) stretched rubber is shown diagrammatically in figure 3. A rather similar picture to 3 (*b*) has been put forward by Mark (1940) to represent the structure of partially crystalline cellulosic fibres, with which rubber possesses many common features. In unstretched crystalline rubber the crystallites are oriented at random. These crystallites do not occupy the whole of the volume; there must necessarily be an appreciable proportion of amorphous rubber, due to the presence of complex entanglements and portions of molecules which cannot be fitted into the available

lattices. In stretched crystalline rubber the forcible extension of the molecules assists the crystallization; the crystallites are all nearly parallel, and the amount of amorphous rubber present is smaller than in unstretched crystalline rubber. From the breadth of the spots in the x-ray diffraction photographs the length of the crystallites has been estimated to be 300–600 A. (Meyer and Mark, 1928); since the average chain length is 20,000 A., it follows that an average molecule must pass alternately through many crystalline and amorphous regions.

The process of crystallization in unstretched rubber is very slow compared with that in a low-molecular compound. According to Bekkedahl (1934) it is most rapid at temperatures between $-35°$ and $-15°$ c., within which limits of temperature the crystallization of raw rubber is completed in a few hours. At $-50°$ c. no crystallization at all was observable. Similar observations have

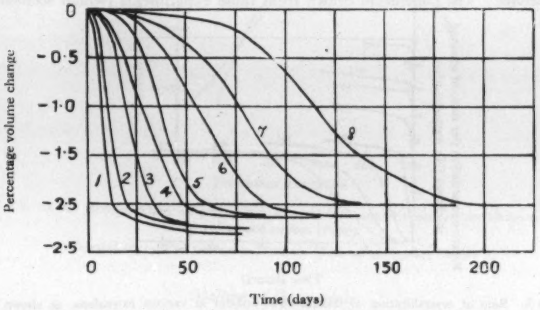


Figure 4. Dependence of rate of crystallization (measured by change of density) of vulcanized rubber on amount of combined sulphur. Temp. 2° c. (Bekkedahl and Wood.)

1. 0·0 and 0·1 % S.	5. 0·40 % S.
2. 0·20 % S.	6. 0·43 % S.
3. 0·30 % S.	7. 0·46 % S.
4. 0·35 % S.	8. 0·50 % S.

(Reproduced, by kind permission, from 'Industrial and Engineering Chemistry'.)

been reported by Meyer and Ferri (1935) and others. The effect of vulcanization is to reduce the rate of crystallization, as shown by the accompanying curves (figure 4) obtained by Bekkedahl and Wood (1941), who made use of the accompanying change in density to study the changes in crystallization. Apparently cross-linking affects the mobility of the molecules sufficiently to impede their "shaking down" into a crystalline lattice, but it affects only slightly the amount of crystallization ultimately arrived at. Furthermore, the temperature of melting of the crystals ($\sim 14°$ c.) appeared to be independent of the vulcanization. Thiessen and Wittstadt (1938) followed the changes in crystallization in stretched vulcanized rubber by observing the double refraction, which was shown to be closely related to the amount of crystallization. This method, as well as the measurement of density, has been used (Treloar, 1941)

to follow the growth of crystallization in raw rubber maintained at 0° c. at extensions ranging from 34 % to 870 %. The optical data, shown in figure 5, reveal a continuous gradation in rate of crystallization from the slightly stretched to the fully stretched rubber, and leave no doubt that the slow crystallization of unstretched rubber and the very rapid crystallization of highly extended rubber are essentially the same process.

The density was found to vary in a very similar manner to the birefringence, the final increase of density ranging from 2·24 % for the unstretched rubber to 3·06 % for rubber stretched to 700 % elongation before freezing. For extensions from 100 % upwards the maximum change in density was approximately proportional to the birefringence; below 100 % extension the birefringence was relatively lower, due to the imperfect orientation of the crystallites at low extensions. The conclusion drawn from these experiments (which seemed to

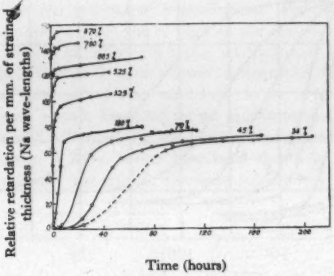


Figure 5. Rate of crystallization of unvulcanized rubber at various extensions, as shown by changes in birefringence. Temp. 0° c. (Treloar.)

(*Reproduced, by kind permission, from the ' Transactions of the Faradcy Society '.*)

be confirmed by Gehman and Field's (1939) x-ray experiments) was that a comparatively small elongation leads to a relatively high degree of orientation of the crystallites in stretched rubber. By extrapolation of the density data it was inferred (Treloar, 1941) that the change in density corresponding to the maximum extension used in the optical experiments was 3·75 %. The density of the rubber crystal must therefore exceed that of the amorphous rubber by about 4·0 %.

Crystal growth curves of the type depicted in figures 4 and 5 are consistent with the view that the growth proceeds from nuclei, "which may consist merely of an approximate parallelism of neighbouring portions of long-chain molecules" (Gehman and Field, 1939).

Relative amounts of crystalline rubber

From x-ray photographs Field (1941 a) has estimated the relative amounts of the crystalline and amorphous components present in stretched rubber. His data, some of which are reproduced in figure 6, show a maximum of about 80 % crystalline material, for both raw and vulcanized rubber at room temperature.

This does not represent the highest degree of crystallization obtainable, and it seems probable, from a comparison of the birefringence data at 0° c. and 25° c. (Treloar, 1941), that at the fullest extension at 0° c. at least 90 % of the rubber is crystalline.

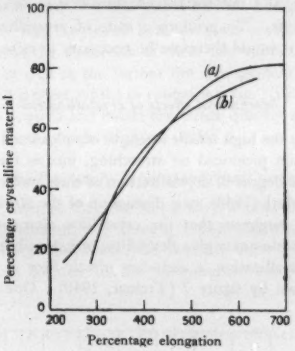


Figure 6. X-ray determination of proportion of crystalline material in (*a*) unvulcanized and (*b*) vulcanized rubber (Field).

(*Reproduced, by kind permission, from the 'Journal of Applied Physics'.*)

Melting of the crystals

In unstretched rubber the crystals melt, not at a sharply defined temperature, but over a range of about 10° c. (Bekkedahl and Wood, 1941 b). The temperature of melting depends on the temperature of crystallization and increases with the time of storing of the rubber over a period of several years. This is another respect in which crystalline rubber differs from crystalline substances of low molecular weight. Similarly the temperature of melting of the crystals in stretched raw rubber depends on the temperature of stretching (Treloar, 1940). In explanation of effects of this kind, Alfrey and Mark (1942) point out that if we consider successive segments of a randomly-kinked molecule attaching themselves to a crystallite, the free energy of the crystallite will increase by a constant amount with each segment added, but the decrease in free energy of the "amorphous" or statistically-kinked portion of the molecule will become progressively less (as the statistical theory shows) with each segment subtracted from its length. The crystallization will proceed to the stage at which the increase in free energy of the crystalline component is balanced by the decrease in free energy of the amorphous component. An increase in temperature will, therefore, not melt the whole crystal but will merely shift the point of equilibrium, causing a greater length of molecule to detach itself from the lattice. A cruder, but perhaps more tangible, way of expressing this conception would be to say that there is always a tendency for the uncrystallized

portions of molecules, being under tension, to pull the crystallized portions away from the lattice. As crystallization proceeds the amorphous portions become more extended, and a point is eventually reached at which the resultant molecular tension balances the molecular attractive forces in the lattice. One imagines that with the lapse of time some of the molecular entanglements are slowly broken down, and the molecular tensions are relaxed, enabling further crystallization to occur. To produce a state of crystallization corresponding with the initial state it would therefore be necessary to raise the temperature.

Mechanical effects of crystallization

It is probable that the high tensile strength of vulcanized rubber is connected with the crystallization produced on stretching, just as the strength of textile fibres is related to the degree of crystallization of their molecules (Harris, Mizell and Fourt, 1942). Mark (1940), in a discussion of the structure and properties of fibrous materials, suggests that the crystalline elements provide strength whilst the amorphous elements give flexibility, or extensibility.

The effect of crystallization in reducing plastic flow in raw rubber is very noticeable, as is shown by figure 7 (Treloar, 1940). One sees from this that

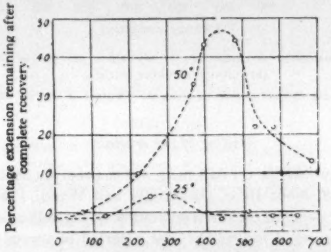


Figure 7. The plastic deformation in raw rubber held for 1 hour at the extensions indicated (Treloar).

(Reproduced, by kind permission, from the 'Transactions of the Faraday Society'.)

the flow increases rapidly until crystallization sets in, when it reaches a maximum and subsequently falls. The entanglement-cohesions postulated as linking the molecules together (§ 4) evidently would not by themselves be sufficiently strong to prevent considerable flow at high extensions; they are just strong enough, however, to maintain the structure until the onset of crystallization, which results in the formation of new and stronger intermolecular linkages.

The converse effect—the increase in toughness of a crystalline body due to the presence of an amorphous high-molecular component—is apparent in a material like polyethylene. Polyethylene crystallizes almost immediately on

cooling, and has comparatively little extensibility. But it contains enough amorphous material linking together the crystallites to make it flexible and tough, in contrast to the chemically similar paraffin wax, which is inflexible and brittle.

Gutta-percha and balata

A brief reference may be made here to gutta-percha and balata. It is generally supposed that the molecules of these two rubbers differ from that of natural rubber only in that in the former the arrangement of the single bonds about the double bond is *trans*, whilst in rubber it is *cis*. This apparently favours crystallization, for both gutta and balata crystallize quickly at room temperature, the crystals melting in the neighbourhood of 60° c.

The β-anomaly in the specific-heat curve

It is satisfactory to note that the so-called β-anomaly in the specific-heat curve of rubber (Ruhemann and Simon, 1928), which showed considerable variation and lack of reproducibility, has recently been shown to be a spurious effect due to crystals formed in the temperature range $-10°$ to $-20°$ c. (Wood, 1942).

§6. THE TRANSITION TO THE GLASSY STATE

Our knowledge of the transition from the elastic to the glassy state rests largely on the work of Bekkedahl and his collaborators. From figure 8, taken from his publications (Bekkedahl, 1934), it is seen that the transition, which is associated with a change in expansion coefficient by a factor of nearly 3,

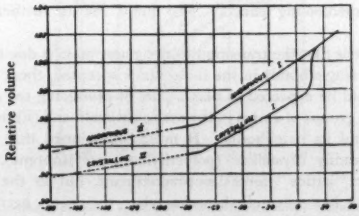


Figure 8. Volume-temperature relationship for purified rubber, showing "second order" transition at $-72°$ c. and melting of crystals at 11° c. (Bekkedahl).

(*Reproduced, by kind permission, from the 'Journal of Research, National Bureau of Standards'.*)

is common to both amorphous and crystalline rubber. The transition temperature ($-70°$ c.) is accompanied by changes in other physical properties, such as dielectric constant (Bekkedahl, 1934), specific heat (Ruhemann and Simon, 1928; Bekkedahl and Matheson, 1935), and thermal conductivity (Schallamach, 1941).

This low-temperature transition, which is of the type known as second-order transitions, presumably corresponds with the onset of some form of molecular vibration. The temperature of the transition is in the same region as the change in amorphous rubber from the rubber-like to the glass-hard state. There can be little doubt that, in general terms, the change in mechanical properties is associated with the loss in freedom of molecular movement; the property of relative freedom of rotation about valence bonds is lost, the atoms becoming relatively fixed in position, as in a glass. It is not possible to say exactly how this loss of molecular freedom of motion is related to the change in other physical properties, such as expansion coefficient. In this connection, the occurrence of the transition in the crystalline rubber presents difficulties. Alfrey and Mark (1941) consider the theory that the transition is due to molecular rotation to be supported by x-ray evidence on the rotation of molecules in paraffin crystals (Müller, 1940), but it should be pointed out that Müller's observations referred to transitions occurring near the melting point. It must be remembered that the crystalline rubber contains a considerable proportion (~30 %) of amorphous rubber, but if the transition effects were due simply to the residual amorphous component, they would be considerably smaller in the "crystalline" than in the amorphous material. Actually, the volume changes are almost identical, and though the discontinuity in the specific-heat curve is sharper in the amorphous than in the "crystalline" rubber (Bekkedahl and Matheson, 1935), the differences are not great. A further difficulty is raised by the observation that the transition in thermal conductivity, when the temperature is falling, may not take place until a temperature 50° c. below the equilibrium transition temperature (Schallamach, 1941). The glassy state is usually regarded as a non-equilibrium state, in which the liquid structure has become "fixed"; it is difficult to believe that the transition to the supercooled liquid state can itself be subject to supercooling effects. The point merits further experimental investigation.

If the hypothesis that the transition to the glassy state is due to the cessation of rotation about single bonds in the molecule is accepted, there are two sets of forces which would be expected to play a part in producing the effect, namely: (i) forces between groups of atoms in the molecule itself, and (ii) forces between a given molecule and its neighbours. It must be assumed that in the case of rubbers which readily crystallize (polyethylene, polyisoprene), the effect of intermolecular or "lattice" forces is predominant, but in the case of more complex molecules, like polymethyl methacrylate, the energy barriers to internal rotation may be comparatively high, and it may be that it is to this internal "stiffening" rather than to the intermolecular forces that the low-temperature transition should, in such cases, be ascribed. It is to this factor that the difference between polymethyl acrylate, which is elastic at room temperature, and polymethyl methacrylate, which becomes rubber-like at about 90° c., has been attributed (Tuckett, 1942). On the other hand, the transition temperature of many polymers which are normally glass-like may be reduced by many tens of degrees by the incorporation of a "plasticizer" (or, more correctly, "elastifier"). This is true, for example, of methyl methacrylate (Tuckett, 1942), and seems to throw doubt on the adequacy of the hypothesis of internal stiffening. The "elastifier"

reduces the forces between molecules, but clearly cannot affect the forces restricting rotation in the individual molecule.

In other cases the transition temperature may be *raised* by the introduction of foreign molecules. This happens when divinyl benzene is incorporated into polystyrene, and has been attributed to its well-known cross-linking effect, which leads to reduced molecular mobility (Ueberreiter, 1940). The increase in transition temperature on vulcanization of rubber by sulphur has been similarly explained (Ueberreiter, 1940). In the latter case it is doubtful whether the cross-linking is the only, or even the most important, cause of the change; the mere addition of sulphur to the molecule, without cross-linking, would increase the inter-molecular forces and hence affect the transition temperature (Gee, 1942).

The transition temperature and relaxation phenomena

Whatever the view taken as to the relative importance of the various factors discussed above in determining the transition from the glassy to the elastic state, there seems to be general agreement with the theory that at this temperature the atoms of the molecular chain cease to be fixed, as in a solid, and acquire a degree of mobility comparable with that of the molecules of a liquid, but limited, of course, by their being joined together as members of a single chain. As in the case of a liquid, the chain atom may be thought of as normally occupying a potential "hole" from which it escapes from time to time on receiving the requisite thermal energy.

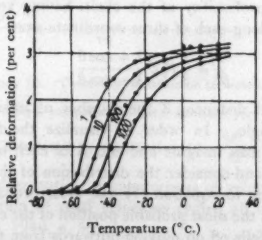


Figure 9. The relation between amplitude of deformation of vulcanized rubber (3% S) and temperature. Parameter:—frequency in cycles per minute. (Aleksandrov and Lazurkin, 1939.)

(*Reproduced, by kind permission, from the 'Journal of Technical Physics, U.S.S.R.'*)

Whether, or not, a substance is highly elastic will depend on whether the atoms have, or have not, sufficient energy to escape from their potential "holes" in the time during which the observation is made. It follows that a substance may be rubber-like when the deformation takes place slowly, but not rubber-like under rapid deformation. This aspect has been studied by Aleksandrov and Lazurkin (1939), who investigated the relation between amplitude of vibration and temperature in rubbers subjected to an alternating compressive force

varying in frequency from 1 to 1000 cycles per minute. Figure 9, which is typical of their results, shows that the transition temperature for vulcanized rubber was raised by about 20° c. when the frequency of vibration was raised from 1 to 1000 cycles per minute. Their curves could be interpreted on the basis of the following equation:

$$D = D_\infty(1 - e^{-t/\tau}),$$

relating the deformation D with the time of application t of a constant force, D_∞ being the deformation when $t = \infty$. In this equation τ appears as a time of relaxation, or "time of orientation", of the molecules, which was related to the potential barrier U restricting the change of form of the molecule by the usual formula,

$$\tau = Ae^{U/kT}.$$

This relation adequately represented the experimental data. It is to be hoped that the evaluation of U for different materials will help in the understanding of the processes involved in molecular orientation and viscous flow in polymers.

§7. KINETIC THEORY APPLIED TO A MOLECULAR NETWORK

We return now to the development of the kinetic theory of elasticity, already briefly considered in §2. The distribution formula derived by Guth and Mark (1934) and by Kuhn (1934) to represent the length distribution of a paraffin chain is

$$p(x, y, z)\,dx\,.\,dy\,.\,dz = \frac{\beta^3}{\pi^{3/2}} e^{-\beta^2(x^2+y^2+z^2)}\,dx\,.\,dy\,.\,dz, \qquad \ldots\ldots(1)$$

where $p(x, y, z)$ is the probability of the chain having components of length x, y and z respectively along each of three coordinate axes. In this equation

$$\frac{1}{\beta^2} = \frac{2}{3} l_c^2 Z \frac{1 + \cos\theta}{1 - \cos\theta}, \qquad \ldots\ldots(1a)$$

l_c being the C—C bond distance, Z the number of links in the chain, and $180° - \theta$ the valence angle. In order to visualize the distribution function represented by (1), one may imagine one end A of each molecule fixed at the origin of coordinates O, and consider the distribution of the coordinates of the other ends B. According to equation (1), the most probable position of B is given by $x = y = z = 0$, i.e. the most probable position of the end B is at the origin. The probability density falls off on moving outwards from O, in the way shown in figure 10 (*a*), representing the function

$$p(x) = \frac{\beta}{\sqrt{\pi}} e^{-\beta^2 x^2}$$

This does not mean that the most probable length of the chain, *when all directions are considered*, is zero. To find this we write $r\,(= (x^2 + y^2 + z^2)^{\frac{1}{2}})$ for the length of the chain and determine the number of molecules for which the end B lies within a volume bounded by spheres of radii r and $r + dr$ respectively. From equation (1) this is (Kuhn, 1934)

$$\chi(r)dr = \frac{\beta^3}{\pi^{3/2}} \cdot 4\pi r^2 e^{-\beta^2 r^2} dr, \qquad \ldots\ldots(2)$$

which is zero when $r=0$ and reaches a maximum when $r=1/\beta$ (figure 10 (*b*)). Equations (1) and (2) correspond to those met with in the kinetic theory of gases,

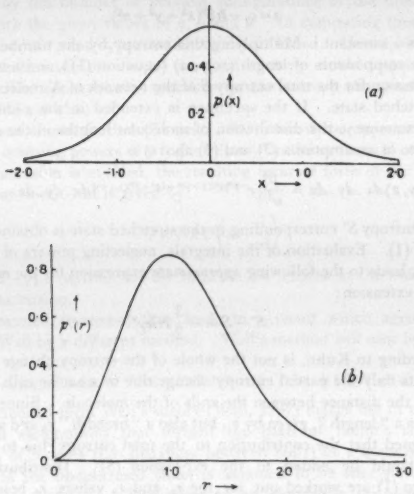


Figure 10. Theoretical distribution of molecular lengths.

(*a*) Length component in fixed direction $p(x)=\frac{\beta}{\sqrt{\pi}}e^{-\beta^2x^2}$.

(*b*) Total length in any direction $p(r)=\frac{4\beta^3}{\sqrt{\pi}}r^2e^{-\beta^2r^2}$.

representing (1) the distribution of components of velocity in a given direction and (2) the distribution of total velocities (when all directions are included) respectively (Roberts, 1940, p. 66).

§8. KUHN'S TREATMENT OF THE NETWORK

In order to simplify the problem of the elasticity of a network of molecules, Kuhn (1936) made the following assumptions:

(i) All the molecules have the same chain-length.

(ii) The distribution of length components of the molecules is that represented by equation (1).

(iii) On extension the components of length of each molecule change in the same ratio as the components of length of the macroscopic rubber.

(iv) The volume of the rubber remains unchanged during deformation.

The following is a summary of Kuhn's treatment. An expression is first derived for the entropy s of a single molecule in terms of its length components. This is obtained from Boltzmann's relation $S = k \ln p$, so that, from equation (1). we have for this entropy

$$s = c_1 - k\beta^2(x^2 + y^2 + z^2), \qquad \ldots\ldots(3)$$

where c_1 is a constant. Multiplying this entropy by the number of molecules having the components of length (x, y, z) (equation (1)), and integrating, leads to an expression for the total entropy S of the network of N molecules per c.c. in the unstretched state. If the specimen is extended in the x-direction by the fractional amount γ, the distribution of molecular lengths in the stretched state (making use of assumptions (3) and (4) above) is

$$p'(x, y, z)\,dx\,.\,dy\,.\,dz = \frac{\beta^3}{\pi^{3/2}} e^{-\beta^2[x^2/(1+\gamma)^2 + (y^2+z^2)(1+\gamma)]}\,dx\,.\,dy\,.\,dz. \qquad \ldots\ldots(4)$$

The total entropy S' corresponding to the stretched state is obtained by using (4) in place of (1). Evaluation of the integrals, neglecting powers of γ higher than the second, leads to the following *approximate* expression for the entropy change due to the extension:

$$S' - S = -\frac{3}{2} N k \gamma^2. \qquad \ldots\ldots(5)$$

This, according to Kuhn, is not the whole of the entropy change on extension. It represents only the partial entropy change due to what he calls the r_1 values, where r_1 is the distance between the ends of the molecule. Since the molecule not only has a "length", given by r_1, but also a "breadth" r_2 and a "thickness" r_3, it is argued that the contribution to the total entropy due to these r_2 and r_3 values should be added to the expression (5). Distribution functions analogous to (1) are worked out for the r_2 and r_3 values, r_2 being defined as the distance of the middle link of the chain from the line joining its ends, and r_3 as the distance of links one quarter of the way from each end of the chain (measured along the chain) from the plane containing r_1 and r_2. The result is an additional entropy change on extension amounting to $-Nk\gamma^2$ for each of the r_2 and r_3 values, so that (5) becomes

$$S' - S = -\frac{7}{2} N k \gamma^2, \qquad \ldots\ldots(5\,a)$$

which, on application of the thermodynamic relation between tension F and entropy (cf. Treloar, 1942), i.e.,

$$F = \left(\frac{\partial W}{\partial l}\right)_T = -T\left(\frac{\partial S}{\partial \gamma}\right)_T, \qquad \ldots\ldots(6)$$

gives the linear stress-strain relation

$$F = 7NkT\gamma. \qquad \ldots\ldots(7)$$

§9. AMENDMENT OF KUHN'S TREATMENT

As it stands, Kuhn's treatment is open to criticism. From a study of Wall's work (to be considered later) the author (Treloar, 1943) has been led to argue that the inclusion of the r_2 and r_3 values in the expression for the entropy

change on extension is incorrect. For, in considering the extension of the molecule, one has to consider one end fixed, whilst the other is confined to the element of volume $dx.dy.dz$. The entropy is related to the probability that the end of the molecule shall be in this element of volume, which in turn is determined by the number of possible configurations of the molecular chain consistent with the given values of x, y and z. In computing these configurations all values of r_2 and r_3 are automatically included. Hence, to attribute to them a separate contribution to the total entropy would appear to be unwarranted. The approximate expression for the entropy change should therefore be (5) and not (5 *a*).

It has been shown further (Treloar, 1943) that if the approximation made by Kuhn in omitting powers of γ higher than γ^2 in the evaluation of the entropy change on extension is avoided, the resulting *accurate* form of the stress-strain relation is not linear, but is given by

$$F = NkT\left(\alpha - \frac{1}{\alpha^2}\right), \qquad \ldots\ldots(8)$$

where $\alpha(=1+\gamma)$ is written for the ratio of the lengths in the x-direction after and before extension.

This amended treatment thus leads to a result which agrees with that derived by Wall by a different method. Wall's method will now be considered.

§ 10. WALL'S TREATMENT OF THE NETWORK

In its amended form Wall's theory (Wall, 1942) makes use of the same four fundamental assumptions that formed the basis of Kuhn's analysis. The distribution of lengths (i.e. distances between ends) of a system of N_0 equal molecules, in the undeformed state, is assumed to be that given by Kuhn's distribution formula (1). The effect of an elongation (or compression) in which the length is changed in the ratio $\alpha:1$ is assumed to be to alter the components of length of all the molecules in the same ratio as the x-, y- and z-dimensions of the bulk rubber, so that the distribution of molecular lengths corresponding to the deformed state is described by equation (4), when α is substituted for $1+\gamma$. Wall's method consists in calculating the probability P that the distribution represented by (4) should arise spontaneously, given that (1) represents the probability that a particular molecule has components of length (x, y, z). In the same way he determines the probability P_0 of the most probable (equilibrium) distribution. He thus obtains the relative probabilities of the stretched and unstretched state, i.e.,

$$\ln P/P_0 = -\frac{N_0}{2}(\alpha^2 + 2/\alpha - 3).$$

The entropy change accompanying the deformation is readily obtained from this relation by writing $S = k \ln P$. The final stress-strain relationship thus becomes (for a specimen of original cross-sectional area 1 cm^2)

$$F = NkT\left(\alpha - \frac{1}{\alpha^2}\right) = \frac{\rho RT}{M}\left(\alpha - \frac{1}{\alpha^2}\right). \qquad \ldots\ldots(9)$$

In this equation, which is identical with the amended expression derived by Kuhn's method, F is the force per cm.^{2}, referred to the unstrained section, ρ the density and M the molecular weight. The molecular weight which this represents is considered to be the weight between junction points of the network. It is pointed out that the formula (9) applies equally to extension and to unidirectional compression. The stress-strain curve represented by this equation is shown in figure 11.

The shear deformation has also been dealt with by Wall (1942) by a method differing only in detail from that just described. If σ is the amount of the shear, the result obtained for the strain energy W is

$$W = \tfrac{1}{2}NkT\sigma^2 = \tfrac{1}{2}G\sigma^2. \qquad \ldots\ldots(10)$$

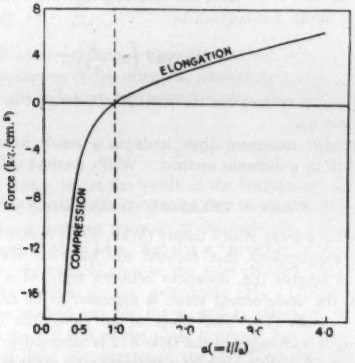


Figure 11. The theoretical stress-strain relationship for a molecular network. $M=15{,}000$, $T=298°$ K.

where G is the modulus of rigidity. Under shear, therefore, the stress-strain relation is linear, whereas under elongation or unilateral compression it is non-linear. The expression (10) may also be derived by application of Kuhn's original method (Treloar, 1943).

Wall's method of approach is superior to Kuhn's in that it does not involve the somewhat unsatisfactory assumption that an entropy value may be associated with the individual molecule * (cf. equation (3)).

It is noteworthy that equations (9) and (10) contain only one molecular constant, namely M, the molecular weight between junction points. The formula should therefore apply not only to a paraffin chain network, but also to any rubber in which the distribution of molecular lengths is represented by a formula of the Kuhn type, e.g. to natural rubber, in which not all the carbon-carbon bonds permit of rotation. But it is important to remember the

* Reference will be made to this difficulty in a later paragraph.

limitations involved in the initial assumptions. In the first place, Kuhn's probability relation (1) applies only so long as the molecule is not extended nearly to its outstretched length. The equations (9) and (10), based on this relation, cannot be expected to hold if the deformation (elongation, compression or shear) is so large that an important proportion of the molecules is fully, or nearly fully, extended. Secondly, the assumption that all the molecules have the same chain length is obviously not even approximately fulfilled. If there is a distribution of chain lengths, the shorter chains will become fully extended at comparatively low extensions. This will limit still further the range of applicability of Wall's equations. In spite of these difficulties, however, equation (9) does give a reasonably satisfactory description of the stress-stain relations in vulcanized rubber (see below).

§ 11. OTHER ATTEMPTS TO SOLVE THE NETWORK PROBLEM

The methods of Kuhn and Wall are not the only ones which have been applied to the problem of the elastic network. They have been given first, and in some detail, because they exemplify in the clearest manner the general principles involved and the type of difficulty encountered.

Pelzer (1938) points out that, if one takes the entropy of the single molecule as a function of the x-coordinate only (corresponding to an extension in the x-direction), by application of the formula $S=k\ln p$ to equation (1), one may derive a corresponding tension f, given by the formula

$$f=2kT\beta^2x,$$

which shows that the tension vanishes only when $x=0$. This clearly cannot be applied directly to bulk rubber. Pelzer therefore assumes that the proper basis for calculating the tension is the distribution function for the r-values (equation (2)). This gives, for the tension f in the molecule,

$$f=-kT\frac{d}{dr}(\ln p(r))=kT(2\beta^2r-2/r).$$

The tension therefore vanishes at a value of r given by $r_0=1/\beta$. Writing $r/r_0=\alpha$, the stress-strain relation *for the individual molecule* then becomes

$$f=2\beta kT\left(\alpha-\frac{1}{\alpha}\right).$$

It is assumed that the stress-strain relation for the bulk rubber is simply Nf, where N is the number of molecules.

The stress-strain curve thus derived does not differ greatly in shape from Wall's (equation (9)), but the treatment is clearly unsatisfactory and arbitrary. It leads to the erroneous conception that the molecule, when held at a length r less than its most probable length $(1/\beta)$, exerts a *compressive* force in the direction of its length, the magnitude of which becomes infinite when $r=0$. Since, according to equation (1), if the end A of the molecule is placed at the origin, the probability density for the other end B is a maximum when B coincides with A, it is clear that the only force which may legitimately be postulated between the ends of a single molecule is a *tensile* force, vanishing when $r=0$.

If at a given instant B happens to be at a distance r less than the most probable distance r_0, then at a later time, though it will most probably be at a greater distance from the origin, it will not be in any preferred direction, and hence cannot be regarded as having moved radially outwards, as if acted on by a repulsive force from the centre. Equation (2), therefore, does not appear to provide a basis for the calculation of the entropy of a single molecule. Indeed, the conception of the entropy of a single molecule seems to involve so many difficulties that it is safer to avoid it altogether.

Guth and James (1941) also considered the problem of the network of molecular chains, making use of the assumption that the volume remains unchanged on extension. Writing L_x, L_y and L_z for the lengths of the sides of a specimen originally in the form of a 1-cm. cube, this assumption means that

$$L_y = L_z = \sqrt{1/L_x}.$$

It is then stated that the probability that the piece of rubber has the dimensions L_x, L_y and L_z is (in a slightly modified notation)

$$p \,.\, dL_x \,.\, dL_y \,.\, dL_z = Ae^{-\beta^2(L_x^2+L_y^2+L_z^2)} dL_x \,.\, dL_y \,.\, dL_z, \qquad \ldots\ldots(11)$$

which, on substitution of the above relation between L_x, L_y and L_z, and multiplication by the number of molecules N, leads to the stress-strain relation

$$F = 2NkT\beta^2(L_x - 1/L_x^2). \qquad \ldots\ldots(12)$$

Guth and James thus anticipate the general form of Wall's equation (9). Their method is, however, unsatisfactory in that equation (11) represents the probability of the molecule having length components L_x, L_y and L_z, and the justification for using it to represent the probability of the bulk rubber having these lengths is not explained.

§ 12. COMPARISON OF THEORY WITH EXPERIMENT

It has been shown (Guth and James, 1941) that the theoretical equation (9) is in reasonably good agreement with the experimental data of Meyer and Ferri (1935) for the elongation, and of Sheppard and Clapson (1932) for the uni-directional compression of rubber, but it is pointed out that the extension and compression branches of the curve thus derived require different values of M. This is to be expected with rubbers differently processed. However, Sheppard and Clapson have also given data for the elongation of rubbers from the same batch as those used in their compression tests, and it would seem preferable to use both their extension and their compression data for the comparison with the theoretical stress-strain relation. The relevant data for such a comparison are represented in figure 12, in which log F (compressive or tensile force in kg. per cm^2 of original section) is plotted against log α (a logarithmic plot is necessary on account of the very large range of F). For large extensions or compressions equation (9) approximates to the respective forms $F = \frac{\rho RT}{M}\alpha$ and $-F = \frac{\rho RT}{M} \cdot \frac{1}{\alpha^2}$, represented in figure 12 by straight lines of slope 1 and -2. The intercepts A, A' of these lines on the log F axis give the values of $\frac{\rho RT}{M}$. It is

seen that Wall's equation gives a very fair approximation to the experimental data, especially for the compression branch, but it is still necessary to use different values of M, i.e. 13,000 and 8,400, to represent the respective branches. The actual values of M are of the expected order of magnitude, but it is difficult to suggest why the effective molecular weight in extension should be 50 % higher than in compression. Nevertheless, to obtain even this extent of agreement over a range in which α varies by a factor of 300 is not unsatisfactory, in view of the very general character of the theoretical assumptions.

It should be observed that the compression data of Sheppard and Clapson were obtained, not by directly compressing the rubber, but by a two-dimensional extension obtained by the inflation of balloons. An indirect method of this kind is necessitated by the fact that it is only possible to compress a rubber block by a small amount (e.g. 50 %) before serious trouble arises from the bulging

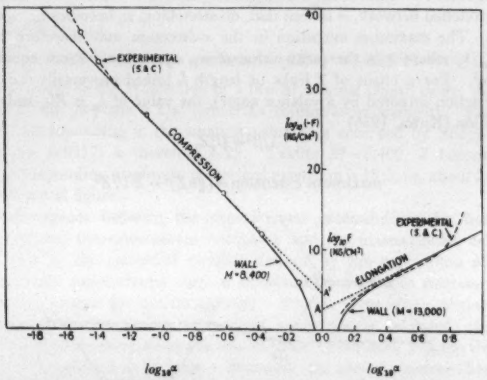


Figure 12. Comparison of theoretical stress-strain relationship with Sheppard and Clapson's experimental data for vulcanized rubber in elongation and compression.

of the block into a barrel-like form as a result of end friction. Provided that the volume remains unchanged by the deformation, however, the two-dimensional extension in the yz-plane may be considered to be equivalent to a unilateral compression in the x-direction, and a very much greater reduction of the x-dimension (to 1/40th of its original value) is thus obtainable.

The significant rise in F above the theoretical curve at elongations or compressions near the breaking point is to be attributed to the pulling-out of the molecules nearly to their fullest extent, an effect not taken into account in the theoretical treatment. From the very high crystallization observed in rubber stretched in one direction it is to be inferred that in the fully extended condition the molecules lie almost wholly in the direction of extension. Likewise, in the compression, or two-dimensional extension of rubber, it is reasonable to suppose that at the point of breaking nearly all the molecules are fully extended in the

plane of the extension. Thus, if breaking did not intervene, each of the two branches of the stress-strain curve would become asymptotic to a vertical line, representing the limit of extension of the network. Although this condition is not reached, it is clearly being approached, and it is possible, assuming as an approximation that the breaking points represent the full extensions of the network, to compare these extensions with the theoretical maximum extensions.

§ 13. MAXIMUM EXTENSIBILITY OF MOLECULAR NETWORK

(i) *One-dimensional extension*

Assuming that in the fully-stretched network each molecule has its greatest possible length l_m, the extensibility may readily be calculated. For if $x_1, x_2, x_3 \ldots$ are the numerical values of the x-components of length of individual molecules in the unstretched network, it is seen that, on stretching, x_1 becomes l_m, x_2 becomes l_m, etc. The maximum extension in the x-direction may therefore be taken to be $l_m/\bar{x}$, where $\bar{x}$ is the mean value of $x_1, x_2 \ldots$, etc. From equation (1), $\bar{x} = 1/\beta\pi^{\frac{1}{2}}$. For a chain of Z links of length l_c joined universally (i.e. without the restriction imposed by a valence angle), the value of l_m is Zl_c, and we have the relation (Kuhn, 1936)

$$1/\beta^2 = \tfrac{2}{3} l_c^2 Z,$$

whence

$$\text{maximum extension} = (\tfrac{2}{3}\pi Z)^{\frac{1}{2}} = 2{\cdot}17 Z^{\frac{1}{2}}.$$

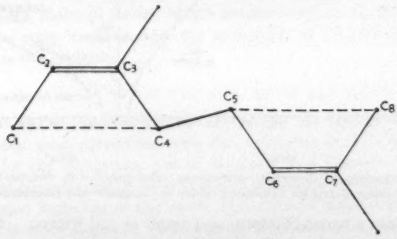


Figure 13. Segment of polyisoprene chain.

The value to be assigned to Z, the effective number of freely rotating links in the rubber chain, has now to be considered.

Figure 13 represents a segment of the rubber chain containing two isoprene units, C_1—C_8 being the chain carbon atoms. The system $C_1\,C_2\,C_3\,C_4$ is rigid, hence it is equivalent to a link of length 3·09 A. (assuming the usual bond lengths and angles), rotating freely about the axis $C_3\,C_4$. $C_4\,C_5$ has the length 1·54 A., and rotates freely about $C_3\,C_4$. The next long link, $C_5\,C_8$, has the same motions relative to $C_4\,C_5$ that $C_1\,C_4$ has. The relatively short $C_4\,C_5$ link therefore acts approximately as a universal joint between the links $C_1\,C_4$ and $C_5\,C_8$, and the rubber chain therefore approximates rather closely to a system of universally jointed isoprene units. The isoprene unit

has the weight 68 on the molecular-weight scale, hence for $M=13{,}000$ the appropriate value of Z is 13,000/68, or 191, and the maximum uni-directional extension is given by $\alpha=30$. This is about four times the actual extensibility ($\alpha\simeq 7{\cdot}6$).

(ii) *Two-dimensional extension*

By an argument similar to that under § 13 (i), the maximum two-dimensional extensibility may be estimated. If $r_1, r_2 \ldots$ are the projections of the lengths of molecules in the plane (yz) perpendicular to the direction of compression, these all approximate to l_m when the network is fully extended. The maximum extension (measured by the change in length of a line in the yz-plane) is therefore $l_m/\bar{r}$, where $\bar{r}$ is the mean value of $r_1, r_2, \ldots$, etc. For the distribution of r values the appropriate function (analogous to (2)) is

$$p(r)dr = \frac{\beta^2}{\pi} \cdot 2\pi r\, e^{-\beta^2 r^2} dr;$$

whence $\bar{r} = \pi^{\frac{1}{2}}/2\beta$, and

$$\text{maximum linear extension} = (6Z/\pi)^{\frac{1}{2}} = 1{\cdot}38Z^{\frac{1}{2}}.$$

If α is the length in the x-direction of a line of original length unity, the linear extension in the yz-plane is $1/\alpha^{\frac{1}{2}}$ (since the volume is unchanged). The linear extension corresponding to the smallest value of α recorded by Sheppard and Clapson, i.e. 0·0257, is therefore 6·25. Taking $M=8{,}400$, Z becomes 123, and the corresponding maximum theoretical extension is 15·3, i.e. about 2·5 times the experimental figure.

The divergence between the experimental extensibilities for both one-dimensional and two-dimensional extension, and the extensibilities calculated on the basis of the molecular weights derived by the application of Wall's equation to the experimental data, is considerable. Various reasons may be suggested to account for the discrepancy. Firstly, there is the obvious point that breaking will occur before the network is completely extended; the actual extensibility will therefore always be less than the theoretical, but the difference would not be expected to be large. Secondly, the theory assumes that all the molecules have the same chain length, whereas in practice there must be a wide distribution of chain lengths. It is not easy to see what effect this will have on the extensibility. The third, and probably the most important, source of error lies in the use of Kuhn's formula for the distribution of molecular lengths, which forms the basis of Wall's theory. The derivation of Kuhn's formula involves assumptions which cannot apply accurately to real molecules. They are: (i) that each link of the chain possesses complete freedom of rotation, so that all configurations may be considered to have the same energy; and (ii) that the space occupied by the chain may be neglected, so that one part of the chain offers no obstruction to the motion of any other part. This would be inaccurate even if the atoms had no volume; the additional volume of the atoms renders it still more inaccurate. With regard to (i) it has been pointed out (Huggins, 1939) that the energy barriers restricting rotation about a single bond in a typical case are such that certain positions, spaced at 120° with respect to each other, are most likely to be occupied. If the three energy barriers were of the

same height, the probability of occupation of each position would be the same, and the resultant distribution of configurations would be nearly the same as if the rotation were free (provided that the height of the barriers did not stop rotation altogether). But if the potential barriers were of different magnitude a certain position would be more probable than the others, and a considerable departure from complete randomness might result. Each of the effects (i) and (ii) is likely to produce an increase in the average length of the chain, i.e. a reduction in extensibility. The above analysis of the experimental data suggests that the actual average length of the rubber chain may be as much as 2·5 to 4 times the average length arrived at from the statistical calculation.

REFERENCES

ALEKSANDROV, A. P. and LAZURKIN, Y. S., 1939. *J. Tech. Phys., U.S.S.R.*, **9**, 1249.
ALFREY, T. and MARK, H., 1941. *Rubber Chem. Tech.* **14**, 525.
ALFREY, T. and MARK, H., 1942. *J. Phys. Chem.* **46**, 112.
BEKKEDAHL, N., 1934. *Bur. Stand. J. Res., Wash.*, **13**, 410.
BEKKEDAHL, N. and MATHESON, H., 1935. *Bur. Stand. J. Res., Wash.*, **15**, 503.
BEKKEDAHL, N. and WOOD, L. A., 1941 a. *Industr. Engng. Chem.* **33**, 381.
BEKKEDAHL, N. and WOOD, L. A., 1941 b. *J. Chem. Phys.* **9**, 193.
BUSSE, W. F., 1932. *J. Phys. Chem.* **36**, 2862.
FIELD, J. E., 1941. *J. Appl. Phys.* **12**, 23.
GEE, G., 1940. *Trans. Faraday Soc.* **36**, 1171.
GEE, G., 1942. *Trans. Faraday Soc.* **38**, 319.
GEHMAN, S. D. and FIELD, J. E., 1939. *J. Appl. Phys.* **10**, 564.
GRIFFITH, T. R., 1934. *Canad. J. Res.* **10**, 486.
GUTH, E. and JAMES, H. M., 1941. *Industr. Engng. Chem.* **33**, 624.
GUTH, E. and MARK, H., 1934. *Mh. Chem.* **65**, 93.
HARRIS, M., MIZELL, L. R. and FOURT, L., 1942. *Industr. Engng. Chem.* **34**, 833.
HUGGINS, M. L., 1939. *J. Appl. Phys.* **10**, 700.
KUHN, W., 1934. *Koll. Z.* **67**, 2.
KUHN, W., 1936. *Koll. Z.* **76**, 258.
MACK, E., 1934. *J. Amer. Chem. Soc.* **56**, 2757.
MARK, H., 1940. *J. Phys. Chem.* **44**, 764.
MEYER, K. H., 1936. *Trans. Faraday Soc.* **32**, 148.
MEYER K. H. and FERRI, C., 1935. *Helv. Chim. Acta*, **18**, 570.
MEYER, K. H. and MARK, H., 1928. *Ber. dtsch. Chem. Ges.* **61**, 1939.
MEYER, K. H., v. SUSICH, G. and VALKO, E., 1932. *Koll. Z.* **59**, 208.
MÜLLER, A. 1940. *Proc. Roy. Soc.* A, **174**, 137.
PELZER, H., 1938. *Mh. Chem.* **71**, 444.
PELZER, H, 1939. *Rep. Prog. Phys.* **6**, 330.
ROBERTS, I. K., 1940. *Heat and Thermodynamics*, 3rd Edn. (London : Blackie).
RUHEMANN, M. and SIMON, F., 1928. *Z. phys. Chem.* **138** B, 1.
SCHALLAMACH, A., 1941. *Proc. Phys. Soc.* **53**, 214.
SHEPPARD, J. R. and CLAPSON, W. J., 1932. *Industr. Engng. Chem.* **24**, 782.
THIESSEN. P. A. and WITTSTADT, W., 1938. *Z. phys. Chem.* **41** B, 33.
TRELOAR, L. R. G., 1940. *Trans. Faraday Soc.* **36**, 538.
TRELOAR, L. R. G., 1941. *Trans. Faraday Soc.* **37**, 84.
TRELOAR, L. R. G., 1942. *Trans. Faraday Soc.* **38**, 293.
TRELOAR, L. R. G., 1943. *Trans. Faraday Soc.* (in press).
TUCKETT, R. F., 1942. *Trans. Faraday Soc.* **38**, 310.
UEBERREITER, K., 1940. *Kunststoffe*, **30**, 170.
WALL, F. T., 1942. *J. Chem. Phys.* **10**, 485.
WIEGAND, W. B. and SNYDER, J. W., 1934. *Trans. I.R.I.* **10**, 234.
WOOD, L. A., 1942. *J. Chem. Phys.* **10**, 403.

RECENT RESEARCH ON COUNTER TUBES

By J. D. CRAGGS,*
Research Department, Metropolitan-Vickers Electrical Co., Ltd.

§1. INTRODUCTION

A Geiger-Müller counter usually consists of a cylindrical container of glass, inside which is mounted a tubular sheet-metal cathode and a co-axial anode wire. The counter is connected to a suitable amplifier and recording device, and is energized by the application of a potential difference of, perhaps, 1000 volts between the electrodes.

It is necessary to consider briefly the fundamental counter circuit shown in figure 1. The passage of an ionizing particle (α ray, β ray, neutron, cosmic ray, etc.) or photon (γ ray, x ray or light) will cause, by virtue of the secondary electrons produced in the counter, a momentary discharge to take place, with

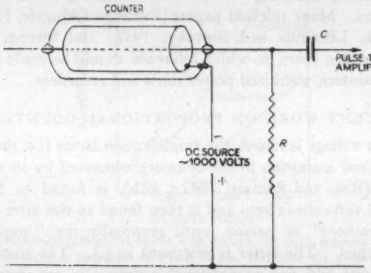


Figure 1. Fundamental counter circuit.

the subsequent development of a voltage pulse on the wire. This pulse after amplification trips the recording device, which is usually an oscillograph or mechanical counter. The Geiger-Müller counter must then return quickly to its original state, in readiness for the next count. The requisite series resistance R and voltage V vary greatly with the type of counter in use. Also, the method of preparation of the counter and the manner in which it is used determine its suitability for a particular purpose. The circuit of figure 1 has been discussed in many publications, but it is only very recently that counter operation has been satisfactorily explained.

* The author is indebted to Dr. A. P. M. Fleming, Director and Manager, Research and Education Departments, Metropolitan-Vickers Electrical Co., Ltd., for permission to publish this *Report*.

The account, by Dr. A. N. May (1938), in Volume v of these *Reports* dealt almost entirely with the mechanism of operation of Geiger-Müller counters, i.e., with the peculiar nature of the transient discharge in the tube. Since that time our knowledge of the modes of operation of counters has been greatly increased, largely by American work, as a glance at the appended bibliography will show. In this connection it is to be regretted that, owing to present circumstances, continental European literature has not been fully reviewed for this *Report*, which deals with the new work on counter mechanisms carried out since 1938. It was considered desirable to include sections on the varied applications of counters, particularly in those branches of physics where their potential uses are wide, but not yet fully exploited, and also on the absolute sensitivity of Geiger counters. In this connection the increasing use of such devices in work on the determination of absolute yields in nuclear reactions may be mentioned.

A great deal of published work carried out before the 1938 *Report* was written was not mentioned therein. Consequently, some references to such literature are included here, where it is considered useful or necessary, although no attempt has been made to compile a complete bibliography. Some important papers, especially for the period 1930–38, have, therefore, not been mentioned.

It is not proposed to discuss the experimental techniques of electrical counting, which have by now become stabilized. A recent book by Lewis (1942) gives details of counting circuits and valuable information on many associated problems. Many relevant papers (Evans and Mugele, 1936; Haines, 1936; Duffendack, Lifschutz and Slawsky, 1937; and Strong, 1939) have appeared in the last ten years, to which reference should be made for constructional details of counters, stabilized power units and recorders.

§2. RECENT WORK ON PROPORTIONAL COUNTERS

As the cylinder voltage is raised, the amplification factor (i.e. the ratio of the pulse charge to initial ionization in the counter), measured by an electrometer-triode technique (Rose and Ramsey, 1942a, 42b), is found to be unity for approximately 100 volts above zero and is then found to rise after the so-called "proportional threshold" is passed, until eventually the "non-proportional threshold" is reached. The latter is discussed in §3. The transition region between proportional and non-proportional counting is not sharply defined, and in this section attention is directed to the former. The operative counter potentials are such that the pulse size (in volts) is proportional to the initial ionization at any one voltage. It is pointed out by Weisz and Ramsey (1942) that, as electron attachment may cause variations in pulse shape, it is desirable to use a gas such as argon where such an effect is absent. For their special purpose, i.e., certain cosmic ray experiments, it was even considered necessary to remove all oxygen from the argon by the use of a glowing tungsten wire.

The constant amplification region corresponds to ionization-chamber operation, when the ions formed by the radiation incident on the counter are not multiplied by collision processes. The amplification factor for the proportional region, where collision multiplication is obtained, may thus be determined directly, if the pulse size at a particular voltage is compared with that obtained at the lowest voltages. Alternatively, the initial ionization may be

determined indirectly. The latter procedure has been successfully adopted by Rose and Korff (1941) in the case of α-particle detection, using the specific ionization data of Holloway and Livingston (1936) and Gurney (1925).

Figure 2 (Rose and Ramsey, 1942b) shows some results for a cylindrical proportional counter obtained with an exact technique involving great sensitivity; the smallest currents involved in work of this kind are of the order of 10^{-14} ampere. It is to be noted that the rise of amplification factor with voltage is greater for A/O_2 than for CH_4 fillings; this matter will be discussed further.

A theoretical treatment of proportional counters has been given by Rose and Korff (1941; see also a résumé of the work by Korff, 1942). In this treatment, amplification of the primary ionization, due to the counted particle or photon, is postulated to result directly from the initiation, by the primary ionization, of an electron avalanche. It is clear, therefore, that a somewhat ill-defined minimum counter voltage will be required for an amplification greater than 1. Rose and Korff then made certain simplifying assumptions, namely:

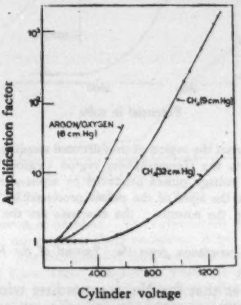


Figure 2. Amplification-factor variation with voltage, for A/O_2 mixture (94% A), and for commercial methane at pressures of 9 and 32 cm. Hg, near the proportional threshold.
(Reproduced, by kind permission, from the 'Physical Review'.)

(i) It is supposed that photo-emission of electrons from the cathode is negligible, and that positive-ion bombardment of the latter also does not give rise to secondary electrons, i.e., that the counter is operated below the Geiger-Müller threshold (see §3). The first part of the assumption was not found to be justified for monatomic and diatomic gases, but is relevant when, for example, CH_4 is used. (ii) Recombination and electron attachment are not encountered. The latter effect should be important only in the presence of, for instance, oxygen and the halogens. Finally, (iii) fluctuations in energy loss and specific ionization are ignored.*

The expression for the amplification factor A, derived by Rose and Korff on the basis of their three assumptions, is

$$A = \exp 2(aN_m c r_1 V_0)^{\frac{1}{2}}[(V_0/V_1)^{\frac{1}{2}} - 1], \qquad \ldots\ldots (1)$$

* The statistical fluctuations in pulse size have been discussed by Weisz and Ramsey (1942).

where a is the rate of increase of ionization cross-section with energy, N_m is the number of molecules of type m per unit volume, $c/2$ is the capacitance per unit length of the counter, the wire of which is of radius r, and V_0 and V_t are respectively the operating and proportional threshold voltages (see also Korff, 1942). The formula is only applicable when the above assumptions hold, e.g., for values of $A<10^4$–10^5. Transition curves showing the change from proportional to non-proportional counting are shown in several papers (C. G. and D. D. Montgomery, 1941 a, 41 b); figure 3 has been obtained by Korff (see Swann, 1940).

Rose and Korff (1941) obtained considerable experimental evidence in favour of their formula for certain gases; the amplification factor was determined from a knowledge of the specific ionization of the Po α particles used in the experiments. Data existed for most of the gases used and the constant for CH_4 was

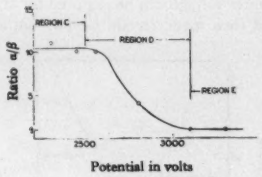


Figure 3. The transition between the region of proportional amplification of the initial ionization in a counter (region C) to the Geiger-Müller region (region E). The ordinates are the ratios of the sizes of the voltage pulses produced by alpha-particles shot into the counter through a thin window to the sizes of the pulses produced by beta particles whose ranges are equal to the length of the counter; the abscissae are the potential differences across the counter.

(*Reproduced, by kind permission, from the 'Journal of the Franklin Institute'.*)

taken as being the same as that for Ne, a procedure which is justified by their similarity in structure. The amplification factor is given by

$$A=cP/en, \qquad \ldots\ldots(2)$$

where c is the capacitance of the anode system, P the pulse size, n the initial ionization and e the electronic charge. The pulse size, in suitable units, may also be expressed (Korff, 1939) as

$$P=(1/c)nf(v)=\sigma lpf(v)e/c, \qquad \ldots\ldots(3)$$

where σ is the specific ionization of the counted particle in the counter gas at pressure p; l is the path length in the tube, and $f(v)$ is the gas amplification at a voltage v. It is clear that a conventional amplifier of finite sensitivity will enable only the upper part of the characteristic to be examined.

It was found (Korff, 1939) that the gases investigated could be separated into two groups. The monatomic and diatomic gases, H_2 and A, and the mixtures, A/H_2, A/O_2 and Ne/H_2, form one group, giving a rapid variation of pulse size with voltage. The second group comprises the polyatomic gases,

and mixtures of polyatomic with other gases, where the proportion of the polyatomic constituent is high (see figure 4). Examples of such mixtures are A/CH_4, A/BF_3, H_2/CH_4 and illuminating gas. These gases are used as fillings in the so-called self-quenching counters, which may be operated with low external resistances (§ 3). The relative performance of the two types of gas is seen in figure 4, taken from Korff's (1939) paper (the A/H_2 counter was necessarily used with a high (10^9 ohms) external resistance, whereas the self-quenching type may be used with 10^6 ohms or less); Stever (1942) has used self-quenching counters with resistances as low as a hundred ohms. It is seen from figure 4 that, in the proportional region below 600 volts, the A/H_2 counter would require a very stable power supply. The A/CCl_4 counter was used with a source emitting α and β rays; the latter were responsible for the rapidly rising part of the curve above 2035 volts, below which only α rays are recorded. Their ability

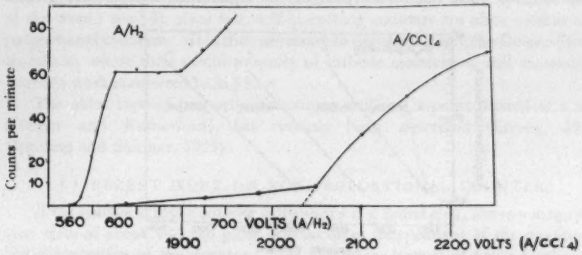


Figure 4. **Characteristics curves of counters.** The curve on the left is that of a non-proportional Geiger-Müller counter, using a quenching resistance of 10^9 ohms. The threshold of this counter is at 575 volts. The curve on the right is for the same counter after CCl_4 has been added to a pressure of 1 cm. Hg. The voltage scales are overlapped but different. The CCl_4 counter is exposed to both alpha and beta rays. Below 2035 volts it counts alpha particles but no beta particles. The right-hand portion of the curve and the dotted extension is the beta-particle response curve. The quenching resistance is 10^6 ohms in this case.
(*Reproduced, by kind permission, from the 'Reviews of Modern Physics'.*)

to discriminate in this way, referred to more fully later, is a most important property of proportional counters.

The essential differences between the two curves of figure 4 have been further emphasized by Rose and Korff (1941). Figure 5, taken from their paper, refers to a copper cathode counter 3 cm. long, 1 cm. in diameter and with a wire of diameter 0·075 mm. and filled with CH_4 and A at a pressure of 10 cm. Hg. It is seen that even in the presence of a large concentration of a self-quenching gas, a monatomic gas still has a considerable effect. The curves, which are extended approximately to the non-proportional threshold, were derived from equation (1), and were fitted to one of the experimental points, since the proportional threshold could not, in these measurements, be detected (cf. figure 2). Stuhlinger (1938), found proportionality in argon-filled counters to extend over a range of approximately 2000 volts. The correctness of Rose and Korff's theory is clearly indicated in figure 5, except for the 50% CH_4 mixture. Similar

curves for a CH_4/A mixture (90% CH_4) at total pressures of 5 to 40 cm. Hg showed that only at the lowest pressure was there disagreement between their theory and experiment. The theory was also successfully applied to the study of A/ether and BF_3/A mixtures, and of the effects of geometrical variations in the counter tubes.

Rose and Korff also put forward an explanation of the difference in amplification of the polyatomic and of the simpler type gases. They point out that the production of ultra-violet radiation by electronic excitation is highly probable only in the monatomic and diatomic gases. Such photons can, in turn, bring about the emission of photo-electrons from the cathode and so cause repeated avalanches, with a consequent rapid rise of amplification factor with voltage (figure 5).*

As striking qualitative support for these hypotheses, Rose and Korff (1941) found that the pulse size for a hydrogen- or argon-filled counter was greater

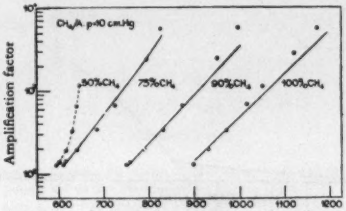


Figure 5. The amplification factor (logarithmic scale) plotted against counter voltage for various relative concentrations in CH_4/A mixtures at 10 cm. Hg pressure. The points are experimental and the full curves are theoretical with one point adjusted. The relative concentration of CH_4 in per cent. is shown for each curve. The dimensions of the counter were: wire diameter 0·075 mm.; Cu cylinder, diameter 1 cm. and length 3 cm.

(Reproduced, by kind permission, from the 'Physical Review'.)

with a non-oxidized sheet copper cathode than with a gauze cathode which effectively gave a 50% reduction in electrode area. Pulse sizes were found to be even smaller with an oxidized gauze cathode. It was, therefore, to be assumed that pulse size increased with increasing photo-electric activity, since existing data indicate that oxidation of copper increases its work function. On the other hand, no such cathode effect was found with counters filled with polyatomic gases, such as CH_4, unless 50% argon was added to the CH_4 (see figure 5) when again the cathode effect was encountered.

It has already been pointed out that self-quenched counters may be operated with a low series resistance ($<10^6$ ohms) in the fundamental circuit, the discharge being quickly completed and quenched (see §3), so that a large circuit time-constant is not necessary for rapid recovery to take place (i.e. for high counting speeds to be obtained). This was realized by Trost (1935 a–37), although the

* *Note added in Proof.* It would now appear that the relative scarcity of ultra-violet radiation, in particular at the cathode, in the fast, i.e. self-quenching, counters is due to the high absorption of such radiation in the complex gases, such as CH_4 (Weisz, 1942 b; Wilkening and Kanne, 1942).

mechanism of self-quenching was not then understood. It is, for instance, often supposed (Korff, 1939) that electron attachment in organic vapours is of importance, especially above the Geiger-Müller threshold (§3). Space-charge effects were invoked by Trost (loc. cit.) and others (May, 1938; Lewis, 1942), but it seems likely that the newer theories outlined above should be included in any explanation of the behaviour of proportional counters. The low mobilities of the complex molecules of, for instance, CH_4 are doubtless important.

The necessity for, and the disadvantages incurred in, using a high resistance to allow of the quenching process in "externally quenched" (an obsolete term) counters is discussed in §3. A discussion of some aspects of the problem has been given by Curran and Petrzílka (1939), and the transition from self-quenched to "externally quenched" counters has briefly been discussed by Weisz (1942a).

Certain important properties of fast counters have been fully described and explained by Stever (1942) in a recent paper. The details are not strictly relevant to a discussion of proportional counters (which are "fast" in the accepted sense of the word; see §3), since fast, self-quenching counters are often used as non-proportional counters. It is then necessary to use them above the Geiger-Müller threshold, where their useful property of cathode inactivity is still maintained. Stever's work is referred to in §3.

The older type of proportional counter, utilizing a point instead of a wire (Geiger and Rutherford), has recently been described (Strong, 1939; Dunning and Skinner, 1935).

§ 3. RECENT WORK ON NON-PROPORTIONAL COUNTERS

If the voltage across a counter is increased it is found that, above a magnification ratio of about 10^8, the pulse size becomes independent of the quantity of initial ionization in the counter. The discharge, instead of being localized in the vicinity of the first electron, spreads along the counter and fills the active portion of the tube. The latter is then said to be operating above the Geiger-Müller or non-proportional threshold (the special definitions introduced by C. G. and D. D. Montgomery are given below), and the well-known plateau is encountered, the extent of which depends partly on the axial symmetry of the tube, the gas filling, and other effects. The latter include such phenomena as the occurrence of spurious counts due to self-excitation of the tube (see below).

The mechanism involved in the transition from proportional to non-proportional counting is complex. The recent important work on positive corona by Loeb and his collaborators (e.g. Trichel, 1939) has not yet been as closely connected with Geiger-counter theory as seems desirable. Weisz (1942a) has discussed the complementary problem of the transition from self-quenching to "slow" counting (see figure 5). The relative lack of photons in self-quenching counters does not seem sufficient in itself to account for their peculiar properties, since the inability of the positive ion sheath, formed near the wire, to promote secondary emission by bombardment on arrival at the cathode must also be explained. The spread of the discharge along the counter, which takes place at the Geiger-Müller threshold, is connected with cathode secondary emission in addition to effects occurring near the wire (Stever, 1942). Photo-ionization in the gas, due to the formation of avalanches, may be responsible for the spreading

of the discharge. Data on the absorption of radiation of wave-lengths around 1000 A., and the relevant collision cross-sections are urgently needed. When operating on the plateau, the counter response is sensibly independent of voltage over an appreciable range of potentials (about 100 volts or more), so that extreme stability of the power supply is not required.

It has already been shown (§ 2) that in the proportional part of the counter characteristic there are two general types of gas filling, namely, those which do and those which do not cause photo-emission of electrons at the cathode, so it is to be expected that these differences will persist in some form above the Geiger-Müller threshold.

The mechanism of non-proportional counting has been studied for some years (Greiner, 1933; von Hippel, 1936), but the most important contributions to the subject appeared in 1940 (Ramsey, 1940 a; C. G. and D. D. Montgomery, 1940 a, 41 a). The Geiger-Müller threshold voltage, i.e., that at which the pulses (~ 1 volt) are all of the same size, is defined (C. G. and D. D. Montgomery, 1940 a, 41 a) as the starting voltage. The difference between the latter and the operating cylinder voltage on the plateau, is the overvoltage. Ramsey's important empirical formula for the wire voltage (V) as a function of time (t) during a counter discharge is

$$V = V_0 - \frac{k}{c} \log \left(\frac{t}{t_0} + 1 \right), \qquad \ldots\ldots(4)$$

where c is the capacity of the counter wire system, k is a constant depending only on V_0, and t_0 is also a constant. The change in V is due to the movement of the positive ion sheath, formed by the avalanche process, away from the wire. The electrons in the avalanche are swept quickly into the wire, but, as they are formed very near the latter where the field is high, their movement does not contribute appreciably to changes in V. Ramsey's formula has also been derived theoretically (C. G. and D. D. Montgomery, 1940 a, 41 a), it being assumed that the total number of positive ions in the space-charge layer, near the wire, increases until the field at the wire is decreased to a value not large enough to produce ionization by collision. The discharge may thus be said to be quenched. The electrons, which, on the average, move through a distance less than the wire diameter, so that their displacement does not contribute appreciably to the change in wire potential, are swept into the wire before the ions have moved appreciably outwards towards the cathode. The space-charge layer then expands, relatively slowly, until the ions reach the cathode. At this stage the difference between slow and fast (self-quenching) counters becomes apparent. In the former case the positive ions cause emission of secondary electrons from the cathode; such secondaries may or may not be multiplied near the wire, depending on the field strength there. In the latter case (fast counters) such a secondary effect is negligible, for reasons which do not yet seem to be clear, and the discharge is complete. Clearly, therefore, for the first case the rate of variation of wire voltage is important; the initial quenching of the discharge due to the presence of the space-charge layer near the wire may thus be vitiated by the existence of a cascaded discharge. The voltage corresponding to the onset of secondary emission at the cathode is near that *usually* defined as the starting potential.

The correctness of this picture is supported by an experiment of Ramsey (1940b), who found that the period of photon emission in a counter was approximately 1 microsecond. The counter wire experienced one-half of its potential change between 1 and 20 microseconds after breakdown, during which period the photon emission was negligible. In this experiment two counters were used in such a way that breakdown in one of them triggered the other counter by a photo-ionization effect. The coincidence rate plotted against the resolving time of the coincidence circuit used gave the required information. Thus any theories that directly connect wire potential changes with photon emission should be reconsidered.

In support of Korff's work on proportional counters (§2) and with reference to the above-mentioned experiment of Ramsey, it is noteworthy that Curran and Strothers (1939) found the counter/counter triggering effect to be much less marked with alcohol/argon (i.e. self-quenching) counters than Greiner (1933) had found with oxygen, hydrogen and air-filled counters.

It has been assumed that the electrons move so rapidly into the wire that their stay in the region between wire and cylinder is of negligible duration. In gases where attachment takes place, the massive negative ions formed may affect the counters' behaviour. The efficiency of an oxygen counter has been determined (Rose and Ramsey, 1941) by the cosmic ray telescope method (Street and Woodward, 1934). For short resolving times the efficiency was 25%, increasing to 80% for longer resolving times. The mean time-lag was 4×10^{-5} second, for the particular voltage condition, and a counter diameter of 2·7 cm.; the probability (q) of electron capture was about 90%, giving a probability (p) of attachment per collision of approximately 10^{-4}. It was shown that

$$q = 1 - \frac{\log\left(\frac{1}{1-\epsilon_0}\right)}{\log\left(\frac{1}{1-\epsilon_\infty}\right)}, \qquad \ldots\ldots(5)$$

where ϵ_0 and ϵ_∞ are the counter efficiencies for short and long resolving times respectively (Rose and Ramsey, 1941). It was also shown that

$$p = 1 - (1-q)^{\frac{1}{N}} \sim \frac{1}{N} \log \frac{1}{1-q}, \qquad \ldots\ldots(5a)$$

where N is the total number of collisions. These formulae have been included here to demonstrate that information on some gaseous collision processes may be obtained by the use of Geiger counters. Rose and Ramsey (1941) also showed that whereas an oxygen-filled counter gave efficiencies varying from 50–80% for resolving times 0·2–70 microseconds respectively, an oxygen/argon counter (6% oxygen, introduced to quench metastable atoms) showed 96% efficiency over the range 0·2–600 microseconds resolving power. The total gas pressure in both cases was 9 cm. Hg, and the cathode was 29 cm. long and 2·7 cm. in diameter. Time-lags in oxygen/argon counters were later studied by C. G. and D. D. Montgomery (1941c).

It has been pointed out (C. G. and D. D. Montgomery, 1940a) that there are two modes of pulse formation. If the wire potential changes rapidly from a

high to a low value so that there is no cathode mechanism,* it will necessarily fall below the starting potential; that is to say, the pulse size is greater than the over-voltage. Thus overshooting (Medicus, 1932) is obtained. The self-quenching, polyatomic, gas mixtures give similar results. Alternatively, as the wire capacitance is made to increase, so the rate of fall of potential and amount of overshooting decrease; the cathode mechanism then becomes operative, since the field at the wire is high enough, when the positive ion sheath reaches the cathode, to cause multiplication of the cathode secondaries, and a series of discharges takes place. The discharge then becomes complex, and occupies a greater time than in the first case, discussed above. The peculiar stepped nature of the slow discharge has been clearly demonstrated by Trost (1935 b, 37) and others (Ramsey, 1940 a; C. G. and D. D. Montgomery, 1940 a).

Evidence has been produced (Ramsey, Hudspeth and Lees, 1941) to show that the quenching mechanism is essentially the same in slow and fast counters, e.g., with argon or ether fillings, but the latter do not give multiple discharges. Ramsey and Lees (1941) have published a short account of experiments with a special slow counter, the cathode of which consisted of nine separate segments, any or all of which could be rendered inoperative, without altering the counter capacitance, by reducing their potentials to suitable values. From the Montgomery theory of counter discharges, assuming a certain cylinder potential and wire capacitance (c), there is a certain minimum counter length (l_e) for overshooting. Shorter counters discharge, multiply, and give a pulse size equal to the over-voltage (about 200 volts pulse, cf. proportional counters, where the pulse is much smaller than this value), whilst longer counters give a single pulse discharge. There is a second order effect due to the finite, but small, time of discharge propagation along the counter, when the length is great and the capacitance small.

For $l > l_e$ (Ramsey and Lees, 1941) the pulse size ΔV is given by

$$\Delta V = \frac{l\alpha}{c}. \qquad \ldots\ldots(6)$$

Here ΔV is the maximum pulse size given by equation (4), and α is the charge per unit length in the space-charge sheath necessary to quench the discharge at a given cylinder potential. The propagation effect changes equation (6) to

$$\Delta V = \frac{l}{c}\left(\alpha - \beta\frac{l}{c}\right), \qquad \ldots\ldots(7)$$

where β is a constant.

For the self-quenching gases the pulse size is proportional to counter length for all values of l, and is inversely proportional to c. (It seems clear that, near the Geiger-Müller threshold the discharge spreads along the wire either on account of cathode secondary emission (C. G. and D. D. Montgomery, 1941 b) or photo-ionization (Stever, 1942) in the gas near the wire.) Ramsey and Hudspeth (1942) have confirmed that with an argon/ether filled (i.e. fast) counter the pulse size is $\frac{l\alpha}{c}$ for all values of c and l.

* i.e. no secondary emission from the cathode due to the positive-ion bombardment.

For fast counters, quenching is caused by the rapid decrease in field strength at the wire, after which the counter recovers by leakage through the series resistance (figure 1). For slow counters, the multiple discharge, if started, continues with decreasing size of ion bursts, until the number of electrons emitted from the cathode, for each burst of positive ions, falls to zero. The discharge then, of course, is extinguished. May (1938, 39) and others (Van Geel and Kerkum, 1938) have investigated the statistics of this process. The function of the series resistance (figure 1) is to allow the wire to recover its original potential after the discharge has been extinguished. For this reason a higher resistance is necessary for slow than for fast counters, because the wire potential must not be allowed to increase appreciably until the multiple discharge is extinguished. R should be kept as low as possible in order to reduce the circuit time-constant and increase the maximum counting rate.

One of the consequences of the new theory (C. G. and D. D. Montgomery, 1940 a) is that a fresh interpretation is put upon the fact that it is desirable to obtain a high value of $i_{\min}$ (the smallest steady current that will flow through the counter in which a glow discharge is maintained). For currents less than $i_{\min}$ the statistical fluctuations in cathode emission are too large to give a maintained discharge. It is clearly desirable that $i_{\min}$, and hence (see equation (8)) $V_{\max}$, should be large, in order that each pulse in the discharge should be large, i.e., that the desirable state of incipient overshooting should easily be attained. Preferably, of course, there should be only a single pulse. On the older theories (Werner, 1934 a, 34 b) $i_{\min}$ was required to be large in order to give a long plateau, since if $V_{\max}$ is the smallest voltage with resistance R that will maintain a discharge and $V_{\min}$ is the voltage corresponding to $i_{\min}$ (i.e., the value towards which $V_{\max}$ tends as $R \to 0$), then

$$V_{\max} - V_{\min} = Ri_{\min} \qquad \ldots\ldots(8)$$

The values of the Geiger-Müller threshold have been calculated for a helium-filled counter by Brown (1942).

If a counter is overshooting it is possible to operate it with a resistance lower than that indicated by equation (8), because the steady state discharge is not developed in the short available time. In other words, the multiple discharge must be avoided if a long counting range is desirable. In a particular case (C. G. and D. D. Montgomery, 1940 a) of an argon/oxygen-filled counter (9 cm. Hg total pressure, 6% oxygen) of cathode length 15 cm. and diameter 1 cm., where $i_{\min}$ was 1·4 microamperes, and the starting potential 785 volts, $V_{\max}$ was found to be 827 volts for $R = 30 \times 10^6$ ohms. The range was, therefore, 42 volts. In practice the counter could be operated over a range of 300 volts for a wire capacity less than 3 cm. Increasing the capacitance reduced the range to about 40 volts, as would be expected from the theory. The upper limit of the range is partly defined by the rate of counting, because a maintained discharge is initiated if a particle passes through the counter when the wire potential, though recovering, is not yet high enough to cause overshooting. It is desirable, therefore, that the starting potential V_s should be small compared with $V_{\max}$. $V_{\min}$ and V_s depend (C. G. and D. D. Montgomery, 1941 b) on the

gas used and, respectively, on the photo-emissive property, and the secondary-emission ratio for positive-ion bombardment of the cathode. Hence it is necessary, in order to obtain as long a range as possible, to increase i_{min} and V_{min} by using a surface of high photo-electric sensitivity but low secondary-emission ratio. It seems that, in the case of self-quenched gases used above the Geiger-Müller threshold (i.e. in the non-proportional region), the secondary emission of electrons from the cathode due to positive-ion bombardment is small.

The upper limit of the range may also be determined by self-excitation of the counter (C. G. and D. D. Montgomery, 1940a, 41b), which gives rise to spurious counts. These seem to be caused by delayed emission of electrons from the cathode due to the presence, on the surface of the latter, of small insulating particles (e.g. oxide, or dust) charged electrically by the action of previous discharges (Malter, 1936). It has been suggested that magnesium or aluminium cathodes should be avoided for this reason, as they are almost invariably oxidized. This is not always borne out in practice (Craggs and Smee, 1942; Bakker, 1937). The presence of metastable atoms may be responsible for the effect (C. G. and D. D. Montgomery, 1941b), but there does not seem to be, contrary to expectation, any definite evidence that counters containing nitrogen or the noble gases are worse in this respect than gases, e.g. hydrogen, where metastable atoms are not found. Oxygen, however, is often used with argon to quench, by collisions of the second kind, the metastable atoms of the latter. Spurious counts due to self-excitation are not distributed at random, because they tend to occur soon after a count has taken place (Medicus, 1932). Also, the number of spurious counts increases more than linearly with the counting rate, due to the fact that a spurious count may cause others to occur. The linear response of the counter to varying intensities of radiation is thus destroyed. This fact may also be demonstrated by increasing the wire capacitance of an overshooting counter. As multiple discharges occur (i.e. as the amount of charge crossing the counter, in unit time, increases), spurious counts commence at an increasingly lower voltage, thus showing a restricted plateau (C. G. and D. D. Montgomery, 1941b). This effect is not found with a self-quenching counter. Spurious counts may be detected either by the methods indicated above or by irradiating a counter and suddenly removing the source. The counting rate, in the absence of spurious counts, will promptly fall to the background level.

The possible maximum rates of counting must now be discussed, i.e. the pulse times for different types of counter. Stever (1942) has fully investigated fast counters with argon/xylol fillings and copper cathodes which were treated to reduce photo-emission. The breakdown time was approximately 10^{-5} second, but this period is followed by a dead-time of about 10^{-4} second, during which the positive-ion sheath has retreated so far from the wire that the field at the latter is equal to the threshold value. A further time, again about 10^{-4} second, is spent before the positive-ion sheath is collected by the cathode (when the wire potential has returned to its operating value), after which the tube is ready for the next count. During this latter period, however, the pulse size is gradually increasing from zero to the correct maximum value. Because of the rapid breakdown, highly resolving coincidence circuits may be used with such counters,

but the maximum rate of counting is limited by the dead time to about 10^4 per second. Stever (1942) has given a theoretical treatment of the movement of the space-charge layer and verified it as closely as could be expected by experimental work. Again, very little information is given on the important subject of secondary emission from the cathode in the presence of polyatomic gases. Some interesting results were obtained with divided cathode and beaded anode counters. In the case of the latter, glass beads were sealed at regular intervals on to the wire. The discharge was found, by observations on pulse size and the dead time, to spread along the counter only when the beads had less than a certain diameter, or if the counter pressure was below a critical value. The discharge (argon/xylol counter filling) did not spread with a divided wire and a single cylinder; this behaviour would be expected on the supposition that the cathode mechanism is inoperative. Further work is necessary for a complete elucidation of the problem, in particular to assess the relative importance, in determining the spread of the discharge, of photo- and electron-ionization effects in the gas occurring in the high field region near the wire.

The clearing time for slow counters depends largely on the time constant (CR* in figure 1), which may be of the order 10^{-3} second. Although the counter discharge, unless overshooting takes place, lasts for a longer time than with a self-quenching counter, it is usually of minor importance. Thus the external circuit largely determines the maximum speed of counting, although, of course, the need for a high resistance is closely connected with the existence of a multiple (i.e. lengthy) discharge in the counter. It is possible to use special circuits (Neher and Harper, 1936; Neher and Pickering, 1938; Lewis, 1942) to increase the speed of quenching in the counter; with these devices R may be reduced to about 1 megohm, and the recovery time to approximately 10^{-4} second.

It is, of course, always necessary in any counting experiment to assess the errors due to the statistics of the count and to the finite resolving power of the counter, counting circuits and recording units. These problems have often been discussed (Locher, 1933; Meixner, 1937; Strong, 1939; Lewis, 1942), and do not warrant further consideration in an article devoted largely to very recent research.

§ 4. THE ABSOLUTE EFFICIENCY OF COUNTERS

In some quantitative experiments, such as yield determinations in nuclear reactions and the determination of the specific ionization of cosmic rays, it is necessary to know the counter efficiencies, the fraction of particles passing through the counter which lead to a count. It is found that nearly every ion pair, produced in the counter, leads to a count, so that the problem is resolved into the determination of the ionization loss of the incident radiation in the cathode and gas filling. For simplicity, the secondary radiation arising in the atmosphere immediately outside the counter and in its glass envelope are ignored. This procedure is justified if all external secondary particles are absorbed in the cathode.

Some data are given in § 5 for counter efficiencies in certain special cases, namely, the counting of light quanta, x rays and neutrons.

* C includes the counter and stray capacitance.

γ rays produce ionization by virtue of the secondary electrons produced in the cathode, through the agency of the photo-electric and Compton effects, and of pair production. Assuming that all the energy of the absorbed photon is transferred to a knocked-on electron, then, approximately, the number of electrons appearing per unit path length is μR, where μ is the absorption coefficient and R the electron range (assumed to be much greater than the counter dimensions) in the counter gas; this product is the counter efficiency. The table below (C. G. and D. D. Montgomery, 1940b) gives some useful theoretical data, the approximate nature of which should be emphasized. Experimental data are not numerous; Read (1935) has published absorption coefficients for x rays (51–21 x.u.) in Pb, Su, Cu and Fe. Additional data for harder radiation are needed.

For instance, the efficiency of a counter with a copper cathode 0·8 mm. thick for 2-Mev. γ rays is calculated to be approximately 0·7%. The agreement with results given below is satisfactory, although it is assumed here that the resolving powers of the counter and circuit are infinite.

Energy (volts)	Absorption coefficient (μ/ρ cm.^{2} gm.$^{-1}$)			Range ($R\rho$ in gm. cm.$^{-2}$)		μR	
	Al	Cu	Pb	Al	Pb	Al	Pb
5×10^5	0·082	0·084	0·121	0·17	0·34	0·014	0·041
1×10^6	0·059	0·057	0·060	0·40	0·80	0·024	0·048
2×10^6	0·041	0·040	0·041	0·89	1·8	0·036	0·074
5×10^6	0·029	0·034	0·043	2·4	4·7	0·068	0·2
1×10^7	0·024	0·033	0·055	5	10	0·12	0·33

Norling (1940) has published some relevant results. For the ThC″ γ rays filtered by 2·5 cm. Pb (i.e. of 2·7-Mev. energy) the efficiencies for Al, Pb and brass cathodes, 0·2 cm. thick, 6 cm. in effective length and 1·4 cm. in internal diameter, were respectively 2·1, 1·9 and 1·7%. The source strength was known from other experiments. At 0·9 Mev., using a ^{56}Mn source with the low-energy β rays filtered out, the efficiencies were about 0·5%. The Pb counter showed a rise in efficiency to 1·1%, due to the increased photo-electric effect, at 0·44 Mev. (^{198}Au). Other theoretical and experimental data were obtained by von Droste (1936), Yukawa and Sakata (1937), and Dunworth (1940). Dunworth's data for a 1·2-mm.-thick brass cathode, which refer to the range 0·2–3 Mev., agree with those of Norling and von Droste. The efficiency of a typical γ-ray counter, with a 0·8-mm.-thick copper cathode, counting with a Ra source filtered by 0·2 mm. of monel metal, was found to be approximately 0·5% (Craggs, 1941), due attention being paid to geometrical factors. The extrapolation of these results to higher energies (in the region of 10 Mev.) is difficult, and further work is necessary before single counters can become accurate quantitative detectors of such high-energy radiations, which arise in many nuclear reaction processes. The low efficiency of γ rays may be increased by the use of a gauze

cathode (Craggs and Smee, 1942; Evans and Mugele, 1936), by allowing the radiation to strike the cathode obliquely, or by using multiple counters (Friedman, Kaiser and Christenson, 1942).

For the experiments described above, it was usually necessary to know the source strength. It is clearly advantageous to eliminate this requirement, and it is at once apparent that this can be achieved by the application of a method similar to that described by Rutherford, Chadwick and Ellis (1930) for the elimination of the personal equation in fluorescent-screen α-particle experiments. Thus, if the counting rate for two tubes, counting separately the same flux of radiation at the same time, are determined, the efficiency of each can be calculated if the separate counts and the coincident count are known. Street and Woodward (1934) have performed experiments with a cosmic-ray telescope arrangement of three counters (see also Kolhörster and Weber, 1941). The efficiency of the central counter was determined by observing the triple-coincidence rate, and the double-coincidence rate for the two outer counters. If corrections are made for accidental counts and showers, it is clearly possible to determine the fraction of particles which passes through the central counter and causes counts. The efficiency was found to be approximately 94%. This result was confirmed (Street and Woodward, 1934) by a second method, which involved the use of separate irradiation of the counters whilst counting coincidences, so that the effect of the inactive period (§ 3) could be obtained directly. Recent experiments of this type are described in detail by Dunworth (1940), who, in collaboration with Feather, has obtained, in this way, much information about nuclear-level schemes (Feather and Dunworth, 1938).

Recently a comprehensive paper by Greisen and Nereson (1942) has appeared, in which experiments, similar to those of Street and Woodward (1939), with argon/alcohol self-quenched counters, are described. The variations in sensitivity along and across the counters have been determined by these authors, it being necessary to determine such quantities for use in experiments where the effective area of a counter must be known (e.g. in work on cosmic-ray showers). If I is the specific ionization of a cosmic ray at atmospheric pressure, p the counter pressure in atmospheres, and l the path length of the ray in the counter, then the efficiency of the counter is $1-e^{-Ilp}$ (C. G. and D. D. Montgomery, 1941 b), or Ilp for low efficiencies. The counter diameter may be put equal to l as a first approximation. Danforth and Ramsey (1936) and Cosyns (1936) have calculated I for certain special cases. These authors found that I(cm.$^{-1}$) for H_2, air, He and A was respectively 5·9, 21, 5·9 and 29·4. For 90% efficiency, when $l=1$ cm., the following gas pressures (cm. Hg) are required: H_2 or He, 29; air, 8·4; A, 5·9. Low-pressure counters may be used in cosmic-ray work to distinguish between single particles and showers, or as discriminating counters to detect, for example, photons in the presence of a background of electrons (Ramsey and Danforth, 1937).

The efficiency of single counters for β-ray detection may, in some cases, be directly measured. Such direct measurements have been made, for example, by Bramley and Brewer (1938) by spreading a sample of U_3O_8 on the end of a counter, using the normal thin window technique. The β radiations from UX_2 and UX_1, which are emitted in equal numbers from thin layers of uranium

salts, have absorption coefficients in aluminium of 14·4 and 510 respectively, corresponding to energies of 80×10^4 and $4{\cdot}2 \times 10^4$ ev. The counter was found to have an efficiency of 11–12%. Correction for a finite source thickness may be made after Orbán (1931).

Some exact experiments have been carried out by Neary (1940) on the β-ray spectrum of RaE, using Geiger counters.

§5. SOME APPLICATIONS OF GEIGER COUNTERS

The detection of photons and the measurement of their intensity, both in the visible and the near ultra-violet, by counters with suitably prepared cathodes has been successfully established as an experimental technique. Duffendack and Morris (1942) have developed counters for use in spectro-chemical studies and compared Geiger counters with electron multipliers and the human eye. As nearly every photo-electron is counted, the dark count due to thermionic emission from the cathode must be reduced as far as possible, if its effect is at all comparable with the normal background count due to cosmic rays and stray radioactive effects. For the detection of visible light (in the region of 4000 to 7000 A.) the cathode is usually covered with a thin layer of an alkali or alkaline earth metal, and the thermionic emission at room temperature may be appreciable. Kolin (1935) has described a technique for activating counter cathodes with Na, Ca or Mg; see, also, papers by Rajewsky (1931) and Locher (1932). A sodium counter (Kolin, 1935) gave an initial count in darkness, presumably at room temperature, of 200 per minute, which decreased to about 10 per minute on the following day.

This effect may be reduced by cooling the counter, perhaps even with liquid air, although this drastic procedure is often unnecessary. Duffendack and Morriss (1942) used a nickel cathode, for which cooling was unnecessary, sputtered in hydrogen for several hours to attain stability. The counter filling was hydrogen, at 10 cm. Hg pressure. The cathode was 0·75 inch long and 0·5 inch in diameter, with a 0·375-inch diameter hole, for the entry of light, cut in one side. A 3-mil tungsten wire was used. The plateau extended from 800–1200 volts. It was found that over the wavelength range 1800–2600 A. 10^4 quanta were needed to release one photo-electron from the cathode; a tungsten-filament lamp was used in this case to irradiate the counter, which was contained in a quartz bulb for transmission below 2600 A. It may be assumed that every photo-electron from the cathode is counted (Locher and LeGalley, 1934; Werner, 1934 a, 34 b) (see §4). At 2600 A. the tube sensitivity was comparable with that of a photographic plate, and much greater than that of a photo-cell.

The yield of 1 photo-electron per 10^4 quanta probably (Duffendack and Morriss, 1932) represents the true yield at 2200–2300 A. Comparable figures given by Barnes and DuBridge (1936) and Kreuchen (1935 a, 35 b) are 10^3 and 10^5 quanta per electron respectively. Duffendack and Morriss (1942) also give yield figures for an eleven-stage $Ag/Cs_2O/Cs$ electron multiplier. The tube was sensitive to visible light; in one case $1{\cdot}5 \times 10^9$ quanta per second were necessary for detection above the dark count. However, the low background count (of about 3 a minute, or 0·05 a second) of the Geiger-Müller counter

enables it to detect $0{\cdot}05 \times 1{\cdot}5 \times 10^{-4}$ or 750 quanta per second; this is a sensitivity greater by a factor of 5×10^7 than that of the multiplier tube. Another multiplier (Bay, 1941) using a potassium coated cathode is referred to by Duffendack and Morriss (1942). In this case the higher work function enables a lower dark count to be obtained, and about 670 quanta per second were detectable, giving a sensitivity equal to that of the Geiger-Müller counter. Finally (loc. cit.) a liquid-air-cooled multiplier tube ($Ag/Cs_2O/Cs$), for which the dark count amounts to only 1 per minute and the yield to 1 electron for every 130 photons, should detect 4 quanta per second. It was found that Geiger-Müller tubes with alumel or Cu_2O/Cu cathodes and hydrogen fillings were sensitive to visible light, but were unstable.

Geiger-Müller counters can also be used to detect x rays. Huppertsberg (1932) has employed a counter having solid metal walls for this purpose. He found that the photo-electrons originated almost entirely from the walls of the counter, the absorption in the gas being negligible.

Locher and LeGalley (1934) have, on the other hand, employed a counter to measure feeble x-ray intensities, utilizing only the photo-electrons liberated by the x rays in the gas of the counter. Assuming that every photo-electron so liberated would be counted, they were able to calculate the absolute sensitivity of the counter in terms of the geometry of the absorbing gas volume, the energy-distribution of the incident radiation and the known absorption coefficients of this radiation in the gas filling the counter. The incident radiation consisted of a carefully collimated beam of Zr fluorescent K radiation, wavelengths between 0·68 and 0·79 A., and the counter was filled with a mixture of approximately 98·5% krypton and 1·5% xenon at a pressure of 6·31 cm. Hg at 27·35° c. Under these conditions, in which the beam traversed a volume of gas in the counter 1·08 cm. thick and 0·0062 cm^2 in cross-section, the calculated absorption in the counter gas was 3·75%. On the assumption that every photo-electron liberated discharges the counter, this figure represents the absolute efficiency; the multiplication factor between the number of photons entering the counter to the number of discharges is therefore equal to 100/3·75, i.e. 26·67. This calculated efficiency was then used to calibrate the x-ray beam and to determine the minimum number of quanta required to give detectable blackening of x-ray film. The counting rate was observed to change from a background of 68·3 a minute to 281 a minute with the x-ray beam switched on under reproducible conditions of x-ray tube current and voltage. The rate due to x rays alone (213·6 a minute) corresponds to 5696·7 quanta per minute (i.e. $26{\cdot}67 \times 213{\cdot}6$) passing through the counter. This corresponds to a flux of approximately $1{\cdot}50 \times 10^4$ quanta per cm^2 per second. Under similar conditions an x-ray film required 45-seconds exposure just to give visible blackening. This corresponded to about 7×10^5 quanta per cm^2 incident upon the film.

Sullivan (1940) confirmed these experiments and found 4–6% absolute efficiency for a counter at about 0·6 A (i.e. about 30 kv. x rays) and increased this figure approximately threefold (to 15%) by using a Zr cathode whose K-series radiation was emitted under the influence of about 0·7 A. primary radiation. Finally (Eisenstein and Gingrich, 1942a) a direct comparison has been made between the sensitivity of a Geiger counter, an x-ray film and an electron multiplier

when measuring x-ray intensities. The counter was 5–10 times as sensitive as the multiplier tube, whose response varied more rapidly with frequency than that of the counter. The application of a counter, with a suitable traversing device, to the study of the diffraction patterns of liquids showed that complete records could be obtained with a counter in about one-twentieth of the time required with the normal photographic method. Some recent applications of Geiger counters to x-ray techniques have been given by other authors (Harding, Smith and Davey, 1942; Eisenstein and Gingrich, 1942b).

It is proposed to omit details of α-, β-, and γ-ray counting, as the techniques are now well known, and references to such subjects have already been given in this *Report*, but it is appropriate here to describe the recent work on discriminating counters, using proportional tubes (see §2). In some cases non-proportional counters may be used, with suitable techniques. For instance, neutrons may be counted with a tube, operating in the Geiger-Müller, non-proportional, region, and containing a silver cathode. The artificial radioactivity of the latter, when the counter is removed from the vicinity of the neutron source, is utilized to give counts (Locher, 1936).

Figure 3 shows results obtained by Korff with a proportional counter; he has also discussed the efficiency of BF_3 counters when thermal neutrons are being detected. The counter efficiency is given as

$$E = Np\sigma_B C_B l, \qquad \ldots\ldots(9)$$

where N is Loschmidt's number, p the counter pressure, σ_B the capture cross-section of ^{10}B, whose concentration in the BF_3 is C_B, and l is the average path length through the counter. E in one case was found to be approximately 3×10^{-3}. Swann (1933) has discussed the determination of l (see also §4). It is apparent (Korff, 1939) that the density of neutrons, ρ (number per cm^3), is the measured quantity, so long as the cross-section is proportional to $1/v$, since the flux of neutrons $= v\rho$. With a BF_3 counter, neutrons having energies from thermal values up to 10^4 ev. can be counted. Korff's curves for starting potentials are given in figure 6 (Korff, 1942) to emphasize the utility of a BF_3 counter in discriminating between γ rays and neutrons. It is pointed out that a uranium gas may be useful for detecting neutrons, by virtue of the heavy ionization caused in the gas by the fission products. Proportional counters may also be used to determine pulse-size distributions for groups of particles of varying ionization loss, by finding the pulse-size distribution/counter voltage and calibrating the counter with particles of known, constant, ionization loss.

Stever (1941) has recently described a directional Geiger counter. This is necessarily of the fast type, with a cathode specially treated to give low photo-emission. It was found that the spread of the discharge (see §3) along the wire could be restricted by sealing glass beads on to the latter. A particle passing along the counter caused pulses of a size proportional to the number of anode segments traversed. When the circuit was set to accept pulses arising from all the segments excited together, then only particles which had passed through the whole counter were registered. Electrons (β rays) passing axially along the tube were counted, without loss of efficiency, in the presence of very intense radial γ-ray irradiation.

Ramsey (1942) has suggested several interesting applications of divided counters, i.e. those in which several separate cathodes share the same wire. By reducing the voltage of alternate segments below the starting potential, the spread of the discharge along the wire, if a self-quenching gas mixture such as argon/ether be used, may be inhibited. This fact indicates that electrons are probably largely responsible for the spreading mechanism, and not photons, unless the absorption coefficient of the latter in the counter gas is extraordinarily high. Such a divided counter is directional in response, as Stever (1941) pointed out. Ramsey (1942) suggested an ingenious extension of this principle, in which one section of a divided cathode counter is used above the Geiger-Müller threshold to give an indication of the total number of particles in a beam passing along the counter, whilst another section, working in the proportional region,

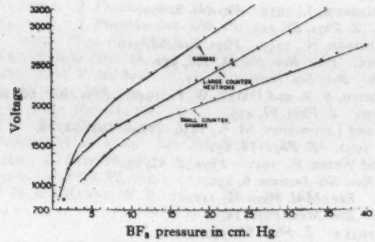


Figure 6. Curves of starting potentials for neutron BF_3 counters, counting neutrons and gamma rays. Note that the voltage difference between the neutron- and beta-detection potentials increases with the BF_3 pressure. Large counter 5·5 cm. diameter, 2·27 cm. long. Small counter 1 cm. diameter, 3 cm. long ; central wire 3 mil. Pulse size constant about 0·1 volt. (*Reproduced, by kind permission, from the 'Reviews of Modern Physics'.*)

responds to the product of the number of particles and their specific ionization. Thus a direct measure of the specific ionization of the particles in the beam is obtained.

One more application of counter tubes may be mentioned. Shive (1939) has described a modulation technique for use with Geiger counters, and a method for determining particle speeds, such as those of thermal neutrons, is outlined. It was found that counters could be modulated, by superimposing an alternating voltage wave on to a constant voltage equal to the counting threshold, at frequencies up to 500 kc. per second. If such a counter were used, 1 metre from a modulated source, and in phase quadrature with it, the speed of a proton or neutron of approximately 10,000 ev. could be measured.

REFERENCES

BAKKER, C. J., 1937. *Physica*, **4**, 723.
BARNES, S. and DUBRIDGE, L., 1936. *Phys. Rev.* **49**, 409.
BAY, Z., 1941. *Rev. Sci. Instrum.* **12**, 127.
BRAMLEY, A. and BREWER, A. K., 1938. *Phys. Rev.* **53**, 502.
BROWN, S. C., 1942. *Phys. Rev.* **62**, 244.

COSYNS, M., 1936. *Bull. Tech. Ass. Ing. Brux.* p. 173.
CRAGGS, J. D., 1941. *Nature, Lond.*, **148**, 661.
CRAGGS, J. D. and SMEE, J. F., 1942. *Brit. J. Radiol.* **15**, 228.
CURRAN, S. C. and PETRZILKA, V., 1939. *Proc. Camb. Phil. Soc.* **35**, 309.
CURRAN, S. C. and STROTHERS, J. E., 1939. *Proc. Camb. Phil. Soc.* **35**, 654.
DANFORTH, W. E. and RAMSEY, W. E., 1936. *Phys. Rev.* **49**, 854.
VON DROSTE, G. F., 1936. *Z. Phys.* **100**, 529.
DUFFENDACK, O. S., LIFSCHUTZ, H. and SLAWSKY, M. M., 1937. *Phys. Rev.* **52**, 1231.
DUFFENDACK, O. S. and MORRISS, W. E., 1942. *J. Opt. Soc. Amer.* **32**, 8.
DUNNING, J. R. and SKINNER, S. M., 1935. *Rev. Sci. Instrum.* **6**, 243.
DUNWORTH, J. V., 1940. *Rev. Sci. Instrum.* **11**, 167.
EISENSTEIN, A. and GINGRICH, N. S., 1942 a. *Phys. Rev.* **61**, 104.
EISENSTEIN, A. and GINGRICH, N. S., 1942 b. *Phys. Rev.* **62**, 296.
EVANS, R. D. and MUGELE, R. A., 1936. *Rev. Sci. Instrum.* **7**, 441.
FEATHER, N. and DUNWORTH, J. V., 1938. *Proc. Roy. Soc.* A, **168**, 566.
FRIEDMAN, H., KAISER, H. F. and CHRISTENSON, A. L., 1942. *J. Amer. Soc. Naval Engng.* **54**, 177.
VAN GEEL, C. and KERKUM, J., 1938. *Physica*, **5**, 609.
GREINER, E., 1933. *Z. Phys.* **81**, 543.
GREISEN, K. and NEREson, N., 1942. *Phys. Rev.* **62**, 316.
GURNEY, R. W., 1925. *Proc. Roy. Soc.* A, **107**, 332.
HAINES, C. L., 1936. *Rev. Sci. Instrum.* **7**, 411.
HARDING, S. W., SMITH, F. R. and DAVEY, W. P., 1942. *Phys. Rev.* **62**, 296.
VON HIPPEL, A., 1936. *Z. Phys.* **97**, 455.
HOLLOWAY, M. G. and LIVINGSTON, M. S., 1936. *Phys. Rev.* **54**, 18.
HUPPERTSBERG, A., 1932. *Z. Phys.* **75**, 231.
KOLHÖRSTER, W. and WEBER, E., 1941. *Phys. Z.* **42**, 13.
KOLIN, A., 1935. *Rev. Sci. Instrum.* **6**, 230.
KORFF, S. A., 1939. *Rev. Mod. Phys.* **11**, 211.
KORFF, S. A., 1942. *Rev. Mod. Phys.* **14**, 1.
KREUCHEN, K. H., 1935 a. *Z. Phys.* **94**, 549.
KREUCHEN, K. H., 1935 b. *Z. Phys.* **97**, 625.
LEWIS, W. B., 1942. *Electrical Counting* (Cambridge: University Press).
LOCHER, G. L., 1932. *Phys. Rev.* **42**, 525.
LOCHER, G. L., 1933. *J. Franklin Inst.* **216**, 553.
LOCHER, G. L., 1936. *Phys. Rev.* **50**, 1099.
LOCHER, G. L. and LEGALLEY, D. P., 1934. *Phys. Rev.* **46**, 1047.
MALTER, L., 1936. *Phys. Rev.* **50**, 48.
MAY, A. N., 1938. *Rep. Prog. Phys.* **5**, 390.
MAY, A. N., 1939. *Proc. Phys. Soc.* **51**, 26.
MEDICUS, G., 1932. *Z. Phys.* **74**, 350.
MEIXNER, J., 1937. *Ann. Phys., Lpz.*, **30**, 665.
MONTGOMERY, C. G. and MONTGOMERY, D. D., 1940 a. *Phys. Rev.* **57**, 1030.
MONTGOMERY, C. G. and MONTGOMERY, D. D., 1940 b. *J. Franklin Inst.* **229**, 585.
MONTGOMERY, C. G. and MONTGOMERY, D. D., 1941 a. *J. Franklin Inst.* **231**, 447.
MONTGOMERY, C. G. and MONTGOMERY, D. D., 1941 b. *J. Franklin Inst.* **231**, 509.
MONTGOMERY, C. G. and MONTGOMERY, D. D., 1941 c. *Phys. Rev.* **59**, 1045.
NEARY, G. J., 1940. *Proc. Roy. Soc.* A, **175**, 71.
NEHER, H. V. and HARPER, W. W., 1936. *Phys. Rev.* **49**, 940.
NEHER, H. V. and PICKERING, W. H., 1938. *Phys. Rev.* **53**, 316.
NORLING, F., 1940. *Phys. Rev.* **58**, 277.
ORBÁN, G., 1931. *Akad. Wiss. Wien. Ber.* **140**, 121.
RAJEWSKY, B., 1931. *Phys. Z.* **32**, 121.
RAMSEY, W. E., 1940 a. *Phys. Rev.* **57**, 1022.
RAMSEY, W. E., 1940 b. *Phys. Rev.* **58**, 476.
RAMSEY, W. E., 1942. *Phys. Rev.* **61**, 96.
RAMSEY, W. E. and DANFORTH, W. E., 1937. *Phys. Rev.* **51**, 1105
RAMSEY, W. E. and HUDSPETH, E., 1942. *Phys. Rev.* **61**, 95.

RAMSEY, W. E., HUDSPETH, E. and LEES, W. L., 1941. *Phys. Rev.* **59**, 685.
RAMSEY, W. E. and LEES, W. L., 1941. *Phys. Rev.* **60**, 411.
READ, J., 1935. *Proc. Roy. Soc.* A. **152**, 402.
ROSE, M. E. and KORFF, S. A., 1941. *Phys. Rev.* **59**, 850.
ROSE, M. E. and RAMSEY, W. E., 1941. *Phys. Rev.* **59**, 616.
ROSE, M. E. and RAMSEY, W. E., 1942 a. *Phys. Rev.* **61**, 198.
ROSE, M. E. and RAMSEY, W. E., 1942 b. *Phys. Rev.* **61**, 504.
RUTHERFORD, E., CHADWICK, J. and ELLIS, C. D., 1930. *Radiations from Radioactive Substances* (Cambridge: University Press), p. 584.
SHIVE, J. N., 1939. *Phys. Rev.* **56**, 579.
STEVER, H. G., 1941. *Phys. Rev.* **59**, 765.
STEVER, H. G., 1942. *Phys. Rev.* **61**, 38.
STREET, J. C. and WOODWARD, R. H., 1934. *Phys. Rev.* **46**, 1029.
STRONG, J., 1939. *Modern Physical Laboratory Practice* (London and Glasgow: Blackie and Son).
STUHLINGER, E., 1938. *Z. Phys.* **108**, 444.
SWANN, W. F. G., 1933. *J. Franklin Inst.* **216**, 559.
SWANN, W. F. G., 1940. *J. Franklin Inst.* **230**, 281.
TRICHEL, G. W., 1939. *Phys. Rev.* **55**, 382.
TROST, A., 1935 a. *Z. tech. Phys.* **16**, 407.
TROST, A., 1935 b. *Phys. Z.* **36**, 801.
TROST, A., 1937. *Z. Phys.* **105**, 399.
WEISZ, P., 1942 a. *Phys. Rev.* **61**, 392.
WEISZ, P., 1942 b. *Phys. Rev.* **62**, 477.
WEISZ, P. and RAMSEY, W. E., 1942. *Rev. Sci. Instrum.* **13**, 258.
WERNER, S., 1934 a. *Z. Phys.* **90**, 384.
WERNER, S., 1934 b. *Z. Phys.* **92**, 705.
WILKENING, M. H. and KANNE, W. R., 1942. *Phys. Rev.* **62**, 534.
YUKAWA, H. and SAKATA, S., 1937. *Sci. Pap. Inst. Phys. Chem. Res., Tokyo*, **31**, 187.

INSOLUBLE MONOLAYERS AT THE AIR-WATER AND OIL-WATER INTERFACES

By A. E. ALEXANDER,
Colloid Science Department, Cambridge

§1. INTRODUCTION

The aim of this *Report* is not to give a detailed and comprehensive account of our present knowledge of this extensive field, but rather to outline certain of the major topics upon which recent interest has been focused. These fall under two broad headings, namely, advances in our general understanding of monomolecular films, or monolayers as they are generally termed, and the application of such studies to other physical and chemical problems.

Although the field covered by the above title has usually been regarded as the domain of the chemist rather than of the physicist, yet much of interest to the latter, and many problems which will need his collaboration for successful solution, can readily be perceived. The fundamental phenomena are now reasonably well established, certainly sufficiently to provide the bio-physicist and bio-chemist with a new approach to certain aspects of their more complex biological systems. This biological application, although one of the most interesting and promising in the whole domain of surface chemistry, must perforce be excluded from this *Report*.

§2. INSOLUBLE MONOLAYERS AT THE AIR-WATER INTERFACE

In his monumental work on thermodynamics Gibbs (1878) referred to the problem of an insoluble surface layer as only "of secondary interest" compared with that of the adsorbed films of soluble compounds. The pioneer researches of Langmuir and Adam, beginning some twenty-five years ago, stimulated widespread interest in these insoluble monolayers, and the decade before the present war witnessed many notable advances and applications to diverse problems. Before outlining some of the recent trends it seems advisable to indicate the principal experimental techniques (of which full details have been given by Adam, 1941).

(i) *Principal experimental techniques*

The earliest, and still the most fundamental, measured property is that of surface-tension reduction, generally termed the "surface pressure" and in conformity with the revised nomenclature denoted here by Π: in most of the published literature this quantity is referred to as F. Π can be measured either directly, by use of a Langmuir trough, or indirectly, by use of such surface-tension methods as the Wilhelmy plate, or the maximum pull on a ring. A known weight of the film-forming material is deposited, generally from solution in a volatile solvent, on the previously cleaned water surface, and the variation in Π with molecular packing found by compressing the film by means of a waxed glass slide.

Accordingly it becomes possible to determine the variation of Π with the area per molecule (denoted by A and usually expressed in A^2/molecule). Certain aspects of these Π-A curves are discussed below, and all that need be said here is that, *in general*, the type of curve obtained is a function principally of the hydrocarbon part of the molecule. Thus, to quote simple examples, stearic acid ($C_{17}H_{35}COOH$) and octadecyl alcohol ($C_{17}H_{35}CH_2OH$) give rather similar Π-A curves, so do the *p*-substituted benzene compounds $C_{16}H_{33}C_6H_4X$, where X may be —OH, —O . CH_3, or —NH_2.

By contrast with the above, the change in phase boundary potential consequent upon the spreading of an insoluble monolayer, and generally termed the "surface potential" (ΔV), is determined by the nature and molecular surface density of the polar or anchoring group(s). It is usual to write $\Delta V = kn$ or $4\pi n\mu$ (by analogy with the Helmholtz equation), where n is the molecular surface density (molecules/cm^2), and k or μ is a constant characterizing the polar group and approximately independent of n.

Several methods have been employed for the determination of the "surface potential", but only the more usual ones will be mentioned here. The form due to Schulman and Rideal (1931) employs an insulated metal needle, coated with polonium to ionize the air-gap, placed close to the liquid surface. A reversible electrode dips into the solution, and by use of a Lindemann or valve electrometer and potentiometer box the change in phase boundary potential due to the monolayer is readily measured. The second method, developed by Yamins and Zisman (1933) and by Porter (1937), employs a vibrating plate placed close to the water surface, and thus acting as one plate of a condenser. The difference in the potentials applied to the vibrating plate, when no current flows, for clean and for film-covered surfaces, is the surface potential (ΔV).

The above Π-A and ΔV-n measurements can give considerable information about the non-polar and polar parts of molecules, as indicated below. Additional insight into the forces involved comes from the study of such mechanical properties of monolayers as viscosity, plasticity and rigidity. Of these, the former has been most studied and, like the "surface potential" above, the "surface viscosity" is expressed as a difference, namely, the change in viscosity of the surface layer due to the spread monolayer. Two principal methods are employed: for films of low viscosity the two-dimensional capillary-flow method, and for viscous and plastic films a method based on the damping of various systems when oscillating in the surface.

The "surface canal" viscometer, due initially to Bresler and Talmud (1933), was developed by Joly and Dervichian (1937), and by Myers and Harkins (1937). The quantity of film passing through the canal in unit time (Q) should, by analogy with the three-dimensional Poiseuille flow, be given by the equation

$$Q = \frac{\Pi d^3}{12\eta l},$$

where d and l are the width and length, respectively, of the canal, and η is the coefficient of surface viscosity. This equation is obeyed only for very narrow widths, Q then increasing less rapidly than calculated as the width is increased. (For detailed discussions see Dervichian and Joly, 1938a–39b; Joly, 1938. 39; Harkins and Kirkwood, 1938a, 38b.)

The damping of suitable oscillating systems, usually a plate or disc attached to a vertical torsion wire, was first used twenty years ago by Wilson and Ries (1923) to examine the nature of films formed on the surface of soap solutions. Subsequently it has been used for insoluble monolayers by Langmuir and Schaefer (1937) and by Fourt and Harkins (1938). The damping, and hence the viscosity, is determined from the logarithmic decrement of successive oscillations, the difference between the clean and film-covered surfaces giving the "surface viscosity". This method, which is of special value for polymeric systems such as synthetic polymers and the proteins, and condensed monolayers, indicates whether viscous flow is Newtonian or anomalous, and is useful in detecting phase changes. (See also Mouquin and Rideal, 1927; Bresler and Bresler, 1940.)

Having thus outlined the principal experimental techniques used to obtain a better understanding of the "two-dimensional world" of monolayers, we can now consider some of the more interesting results which have been obtained.

(ii) *Phase changes in monolayers*

Extensive work upon surface pressures, surface potentials and surface viscosities of different types of molecules has shown that marked analogies exist between these two-dimensional monolayers and three-dimensional matter. The principal two-dimensional phases are shown in table 1 and figures 1 and 2.

Table 1

Phase.	Chief characteristics
I. Gaseous and vapour.	If ideal $\Pi A = kT$, i.e. Π-A curve is rectangular hyperbola.
II. Liquid expanded (Adam) or liquid, L_1 (Harkins).	Liquid of high compressibility, obeying relation $(\Pi - \Pi_0)(A - A_0) = C$.
III. Intermediate or transition.	Liquid of extremely high compressibility. Marked hysteresis usually observed.
IV. Condensed:	
(*a*) Close-packed heads (Adam). Liquid, L_2 (Harkins).	Newtonian liquid of low compressibility. Π-A relationship linear.
(*b*) Close-packed chains (Adam). Solid (Harkins, Dervichian).	Compressibility very low. May show anomalous viscosity, plasticity or rigidity. Π-A relationship linear. Packing approximates to that in crystalline solid.

(a) *Gaseous and vapour films.* By the use of sensitive surface-pressure instruments it has been well established that the ideal relationship $\Pi A = kT$ is approached at areas which are of the order of 1,000–10,000 A^2/molecule for simple long-chain compounds (see Adam, 1941). It is usual to plot ΠA against Π when curves showing an obvious analogy to the PV-P curves of Amagat and others are obtained.

(b) *Liquid-expanded films* (figure 1). A simple theory of the liquid-expanded region has been proposed by Langmuir (1933). On the basis of attractive forces between the long hydrocarbon chains and two-dimensional kinetic agitation of the pola head-groups he derived as the equation of state

$$(\Pi - \Pi_0)(A - A_0) = C,$$

Here Π_0 represents the (negative) spreading pressure of the chains, A_0 is a term to allow for the surface-area occupied by the head-group, and C is a constant (equal to kT for un-ionized groups). This equation has been found to fit the experimental data very closely (Langmuir, 1933; Marsden and Rideal, 1938).

Interesting confirmation of this explanation has been obtained from the recent study of insoluble monolayers at the oil-water interface (Alexander and

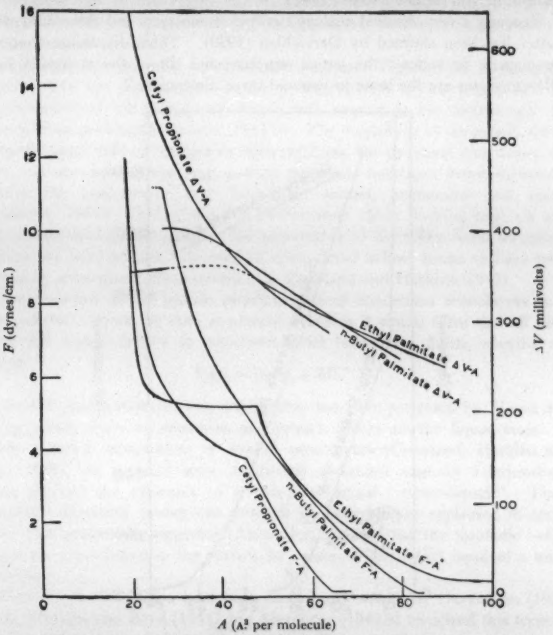


Figure 1. Examples of "liquid-expanded" films, showing transition to condensed region in the case of ethyl palmitate and cetyl propionate (from Alexander and Schulman, 1937).
(NOTE.—Surface pressures are here denoted by F and not by Π.)
(Reproduced by kind permission of the Council of the Royal Society.)

Teorell, 1939; Alexander, 1941 b). Here, as anticipated from the presence of the oil phase, the Π_0 term is eliminated, the observed equation of state being

$$\Pi(A - A_0) = C.$$

(Some further discussion is given below under the oil-water interface.)

(c) *Condensed films* (figure 2). The condensed region, in which the molecular packing closely approaches that in the crystal, has been the subject of

much recent discussion (cf. Dervichian, 1939; Adam, 1941; Harkins and Boyd, 1941; Alexander, 1941c–42b). Adam (1921) early suggested that the two parts XY and YZ (see figure 2) corresponded to close-packed heads and close-packed chains respectively, a view which, with minor modifications and amplification, he still retains (Adam, 1941).

Recently a very detailed analogy between monolayers and three-dimensional matter has been claimed by Dervichian (1939). Thus, discussing condensed monolayers, he states " the lattice structure and tilt of the molecules in the different forms are the same in two and three dimensions "

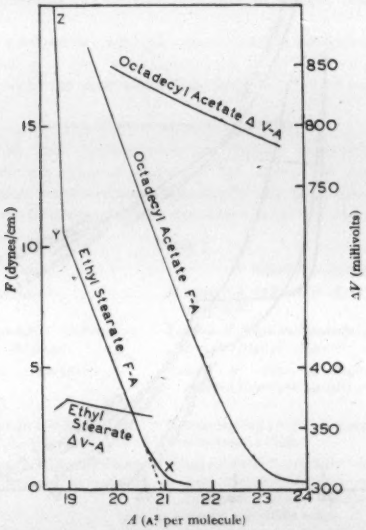


Figure 2. Examples of " condensed " films, showing the more compressible (XY) and less compressible (YZ) regions (from Alexander and Schulman, 1937).

(Note.—Surface pressures are here denoted by F and not by Π.)

(Reproduced by kind permission of the Council of the Royal Society.)

This theory has been criticized by Harkins and Boyd (1941) and by Alexander (1941c), the latter adducing ten arguments against the general validity of Dervichian's thesis. Subsequent work on condensed monolayers, discussed below, has only gone to confirm such criticisms.

Condensed monolayers have been recently approached from two rather different, although eventually complementary, points of view. The first is based upon consideration of thermodynamics and the energetics involved (Dervichian) 1939: Harkins, Young and Boyd, 1940; Harkins and Boyd, 1941).

while the second endeavours to elucidate the structures in terms of precise molecular configurations (Alexander, 1941c–42 b).

For the least compressible region (YZ in figure 2), consideration of evidence based upon monolayer compressibilities, upon surface-potential measurements, and upon the x-ray structure of hydrocarbons near the melting point, suggests very strongly that in this region the long chains are vertically orientated and close-packed, but not quite as tightly as is possible in the three-dimensional crystal (Alexander, 1942b).

With regard to the more compressible condensed region (XY in figure 2), it appears that the configuration at the limiting area X is determined *both* by the orientation of the head-group dipole with respect to the surface and by the long-chain packing (Alexander, 1942 b). The magnitude of the area X seems to depend upon different factors in different cases, the principal ones being the "size" of the head-group (e.g. *p*-alkyl phenols), hydrogen bond formation between the head-groups (e.g. long-chain amides, acetamides and acids) (Alexander, 1941c; 1942a), and the hydrocarbon chain packing (e.g. cis and trans unsaturated compounds). The importance of hydrogen bond formation between the head-groups, discussed in more detail below, seems to have been realized by subsequent investigators (e.g. Copeland and Harkins, 1942).

The question of the precise physical state of condensed monolayers is of great interest. From the data at present available it would seem that all films in the XY region behave as two-dimensional Newtonian fluids, obeying the relation

$$\log \eta = \log \eta_0 + k\Pi.$$

Theoretical justification for this relationship has been provided by Moore and Eyring (1938) from an extension of Eyring's theory of the liquid state. A further detailed examination of alcohol monolayers (Copeland, Harkins and Boyd, 1942) has revealed some apparently abnormal viscosity relationships, which indicate the existence of a two-dimensional "super-liquid". These apparently abnormal results can possibly be very simply explained in terms of the idea previously suggested (Alexander, 1941c), that the alcoholic —OH groups are cross-linked in the surface by means of the —OH bond of a water molecule.

Films in the YZ region have been classed as "solid" by Dervichian (1939) and by Harkins and Boyd (1941), but Alexander (1942 b) criticized this term on account of the obvious fluidity shown by such films as methyl ketones, esters, etc. Anomalous viscosity along YZ is shown by the long-chain acids and alcohols (Fourt and Harkins, 1938; Boyd and Harkins, 1939; Copeland, Harkins and Boyd 1942) and by the long-chain esters (Trapeznikov, 1938–41), and is possibly a general phenomenon for all liquid films in this region; but it would seem better to restrict the term "solid" to those films exhibiting a definite elastic modulus (Trapeznikov, 1938, 39).

(d) *Energetics of phase changes.* The energetics of spreading from the crystal to monolayer, and of phase transformations in monolayers, first studied by Cary and Rideal (1925), have recently been the subject of numerous publications by Harkins and his collaborators. Such determinations involve precise Π-A and Π-T measurements, and a high standard of accuracy has been

achieved. The latent heat of spreading (λ) is found from the Clapeyron equation in the form

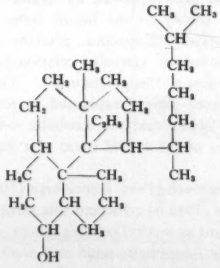
$$\lambda = T(\partial\Pi/\partial T)(A_F - A_S),$$

where A_F and A_S are the areas of surface occupied by film and solid crystals respectively. Harkins and Nutting (1939) thus obtained accurate data upon the latent heat and entropy of spreading of myristic and pentadecylic acids. Harkins, Young and Boyd (1940), Harkins (1941), and Harkins and Copeland (1942) have presented detailed discussions of the thermodynamics of monolayers, such as energy, entropy, and heats of extension and spreading, and the order of phase changes (see also Dervichian, 1939).

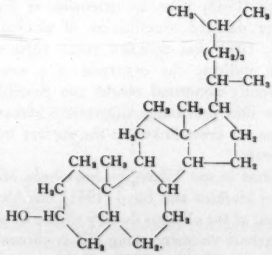
§3. MOLECULAR STRUCTURES FROM MONOLAYER MEASUREMENTS

The determination of molecular structures, particularly of the more complex compounds, now generally involves a co-operative effort between both chemist and physicist. In a number of cases the use of the monolayer technique, particularly in conjunction with x-ray and optical measurements of "built-up" films, as discussed below, has been of value, and it seems well to outline its advantages and limitations.

Thus it is essential that the compound can be made to spread to a homogeneous monolayer, and that, once spread, it is sufficiently stable to enable measurements to be taken. In general, structure determination is restricted to films of



Cholestanol (old formula)



Cholestanol (new formula)

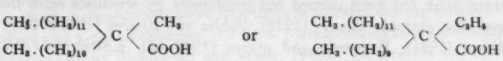
the condensed type, although condensation may, in some cases, be brought about by reduction of temperature, variation of substrate composition, etc. The chief advantage of the technique is the minute amounts required, 5 mgm. being ample for the complete determination of the monolayer and multilayer characteristics. "Built-up" films or multilayers can be used to induce crystallization of otherwise intractable substances, such as unsaturated fatty acids, etc. (see Stenhagen, 1938; Alexander, 1939; Ställberg and Stenhagen, 1942).

The work of Adam and his collaborators (see Adam, 1941) upon surface films

of sterols provides an early example of the use of the monolayer technique in elucidating molecular structure. In the case of cholestanol, for example, consideration of molecular models showed that the older formula required a surface area of not less than 54 A², the new formula one between 35 and 40 A². This can readily be seen from the structural formulae above, if the models are arranged in each case so that the anchoring hydroxyl group is in contact with the aqueous substrate.

Cholestanol gives highly incompressible condensed monolayers, with a limiting area of about 39 A², which thus provides indirect support for the new formulation.

Another excellent example is provided by the recent work of Stenhagen upon phthioic acid, $C_{26}H_{52}O_2$, an optically active acid isolated by Anderson from tuberculosis bacillus. The constitution of this acid is of great importance, as the acid shows almost all the toxic properties of the bacillus itself. Stenhagen determined its monolayer characteristics and compared them with long-chain mono- and di-substituted acetic acids (Stenhagen, 1940), and also made optical and x-ray measurements upon multilayers of the barium salt (Ställberg and Stenhagen, 1942). From these measurements, carried out on a few mgm., he concluded that it was a tri-substituted acetic acid, either



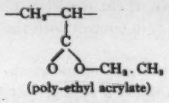
Another major problem still awaiting solution, namely the structure of protein monolayers, which may in its turn throw light on the structure of native proteins, has recently been approached from two fresh angles.

The first was to examine the properties associated with the keto-imino (—CO . NH—) group, the fundamental linkage in the protein molecule, in the presence of an aqueous medium. The part played by hydrogen bonding of the type >C=O———H—N< has been much discussed recently (e.g. Astbury, 1940), and the presence of water seems essential for the proper functioning, if not for the very existence, of the protein molecule.

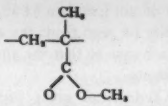
Monolayers of simple long-chain compounds containing this group (e.g. amides, acetamides, ureas, etc.) were accordingly examined, and compared with those of long-chain acetates, esters and methyl ketones, where intermolecular hydrogen bonding similar to >C=O———H—N< is clearly impossible. From the striking differences observed it was concluded that such hydrogen bonding plays a major rôle in condensed monolayers of compounds containing the —CO . NH— group, tending to bring about condensation, solidification, and an increase in half-expansion temperature. Marked sensitivity to substrate *p*H, a well-known characteristic of the protein molecule, was also found. Investigation of other reagents, such as urea, thiocyanates, etc., known to exert marked action on proteins, is clearly indicated.

The second angle of approach has been an examination of monolayers of synthetic linear polymers, such as the poly-acrylates and poly-methacrylates, nylon, etc. (Crisp, 1943). Here, instead of the polypeptide chain —CO . NH . CHR— of the protein, with side-chains (denoted by R) of considerable diversity

and of unknown arrangement along the chain, we have repeating units of known chemical structure, for example,



(poly-ethyl acrylate)



(poly-methyl methacrylate, or "Perspex")

and —CO . NH—$(CH_2)_4$— (nylon), etc. Nylon is of particular interest in view of its close chemical similarity to the polypeptide chain.

§ 4. EVAPORATION THROUGH MONOLAYERS

Illustrating the diversity of problems to which the monolayer technique has been applied, evaporation through monolayers provides abundant theoretical and practical interest. If, by the use of a suitable monolayer, evaporation from lakes and reservoirs could be retarded, a problem of great importance in warmer regions, such as South Africa, Australia and parts of Russia, would have been solved.

Recent work has been carried out principally by scientists from the above countries, for example Glazov (1938), Sebba and Briscoe (1940a–40c), Sebba and Rideal (1941), Heymann and others (1940, 42), Kheinman (1940), and Sklyarenko, Baranaev and Mezhueva (1940). Sebba showed that only condensed monolayers gave appreciable retardation in evaporation rate; with the long-chain alcohols such as cetyl and octadecyl, up to 95% reduction could be obtained at sufficiently high surface pressures. Essentially similar conclusions have been reached by the other investigators in this field. With an alcohol-water substrate the retardations in evaporation of the two components differed, some evidence for selectivity being observed (Sebba and Rideal, 1941).

§ 5. BUILT-UP FILMS OR MULTILAYERS

The interesting observation that certain monolayers on aqueous substrates can be transferred by a simple dipping process to a glass or metal surface was first pointed out by Langmuir (1934). Miss Blodgett, and numerous subsequent workers, have carried out detailed investigations of this phenomenon, the essential points of which are now well established. (For the early references see Langmuir, 1939, and Schulman, 1939.) The changes in interference colour as a multilayer is gradually built up by continued dippings forms a very striking lecture demonstration.

Only condensed monolayers exhibit this phenomenon. The earlier workers used fatty acids, condensed by means of calcium or barium ions in the substrate, but recently numerous other types of compounds (e.g. esters, ketones, phenols, etc.) have been employed.

There are three different types of deposition, *X* when this occurs only on the downward movement of the slide through the monolayer, *Y* on both downward and upward movements, and *Z* on the upward movement only. The monolayer is forced on under the pressure exerted by a "piston-oil", this being a compound having a constant surface pressure.

Bikerman (1939) has shown that the three types of deposition are related to the degree of wettability of the deposited films, and thus to the contact angle. The contact angle depends upon the nature of the compound forming the monolayer, upon the surface pressure, and upon the rate of dipping. As a measure of the wettability he studied the frictional force between a water drop and the multilayer, and thus found that the hydrophilic character of the films increased in the order $X<Y<Z$. Measurement of the wetting angles (θ) between air, water, and the built-up multilayer (Bikerman, 1940) showed that *X*-films were deposited when θ is $>94°$, *Y*-films when θ is $>80°$ and $<92°$, and *Z*-films when it is $<80°$.

One rather interesting point came from the study of the deposition ratio, i.e. the area of slide covered to the area of monolayer removed. This was found to be unity within the experimental error, even for artificially grooved surfaces and wire gauzes (Bikerman, 1939), showing that the deposited monolayer is, at any rate initially, stretched over the undulations just like a soap film would be.

The optical thickness of a multilayer can be determined from its interference colour, polarized light generally being used (for references see Langmuir, 1939). In addition, x-ray or electron-diffraction measurements have been made in some cases upon the same multilayers (e.g. Germer and Storks, 1936–39; Holley and Bernstein, 1937; Stenhagen, 1938; Alexander, 1939; Bernstein, 1940). Such measurements show that there is no exact correspondence between the optical and x-ray thicknesses, the latter corresponding to a known crystalline modification (Knott, Schulman and Wells, 1940). As pointed out above, this technique has much promise for inducing crystallization in some otherwise intractable substances, for example unsaturated acids (Alexander, 1939), and, with synthetic amorphous polymers, Coumoulos (1943) has observed a higher degree of orientation.

An interesting application of "skeleton films", namely to prepare "non-reflecting" glass, has been made by Blodgett (1939). By the use of a suitable substrate a multilayer containing barium stearate with some free stearic acid can be obtained, and on dipping this into benzene the free fatty acid is extracted, leaving a "skeleton" multilayer. This skeleton film has the property of dissolving back hydrocarbons and other compounds, thus enabling films of very different refractive indices to be obtained. By means of a cadmium arachidate skeleton multilayer Blodgett thus produced on a glass surface a film having a refractive index equal to the square root of that of the glass. The intensity of the light reflected from its upper surface is then equal to the intensity of the light reflected from the film-glass boundary after passing through the upper surface of the film. If now the optical thickness of the film be made equal to one-quarter of the wavelength of sodium light, then at normal incidence reflected light of this wavelength is completely extinguished by interference. Treatment of both surfaces in this way makes the glass completely non-reflecting and hence invisible.

Porter and Wyman (1937, 38) found that calcium stearate multilayers of the *X*-type (but not of the *Y*), deposited on a metallic plate, gave contact potentials of about 70 mv. per layer. Thus with a film of 100 layers there exists a potential difference of about 7 volts between a point just above the multilayer and another

point in the air close to the bare metal surface. Langmuir (1938) suggested that these contact potentials were due to recession of the water from the film producing a surface charge on the uppermost monolayer, whereas Goranson and Zisman (1939) ascribe them to absorption of ions from the underlying solution. The latter authors have shown that, if the multilayer is deposited on an insulator, an electrostatic repulsive field is set up which ultimately inhibits the deposition of further layers.

Among the numerous investigations to which these built-up films have been subjected mention may be made of those concerned with their electrical properties, such as dielectric constant and dielectric strength. For the alkaline earth and heavy metal stearates values of between 2·5 and 3·7 (Race and Reynolds, 1939; Zahl, Haskins, Gallagher and Buchwald, 1939) and between 2×10^8 and

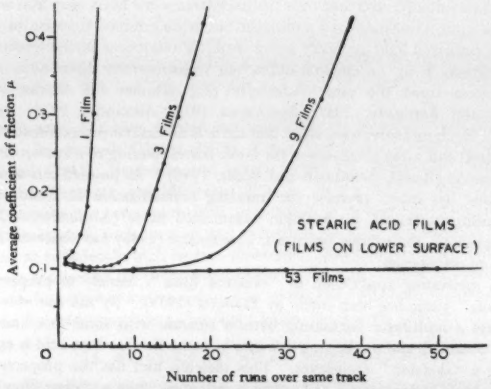


Figure 3. The variation of coefficient of friction with number of runs over the same track for different thicknesses of stearic acid (from Bowden and Leben, 1940).
(*Reproduced, by kind permission, from the 'Philosophical Transactions' of the Royal Society.*)

5×10^8 volts/cm. (Race and Reynolds, 1939; Thiessen, Beischer and von Gillhaussen, 1940) have been recorded for the dielectric constant and the dielectric strength respectively.

Evaporation of free fatty acid from built-up soap films in a high vacuum has been studied by Gregg and Widdowson (1939). A methyl-stearate film containing 700 layers was found by Sutherland and Tutte (1939) to give easily detectable infra-red absorption in the 3·3μ band; weaker bands between 6 and 10μ were also noticeable.

The ability to prepare very thin layers of known thickness by means of the multilayer technique has been turned to good account by Widdowson and Gregg (1939), who used multilayers as absorbing screens in studies of nuclear disintegration.

Another very obvious application of multilayers is to the study of boundary lubrication phenomena (Akamatu and Sameshima, 1936; Bowden and Leben, 1940; Isemura, 1940; Sameshima, Akamatu and Isemura, 1940). With the long-chain fatty acids, for example, a single layer reduces the coefficient of friction (μ) between glass or metal surfaces to one-tenth of its initial value. These very thin layers are, however, easily worn off, and thicker layers are necessary if μ is to be kept at this low value. This fact is clearly illustrated by the results shown in figure 3, taken from the work of Bowden and Leben (1940).

§ 6. APPLICATION OF INSOLUBLE MONOLAYERS TO CERTAIN PROBLEMS OF ADSORBED FILMS

The study of insoluble monolayers has, directly or indirectly, clarified certain much-debated problems of adsorbed films: mention will be made here of the "surface-ageing" phenomenon, the Gibbs adsorption equation, and the structure of surfaces of solutions.

(i) *The surface ageing of solutions*

The anomalously slow rate of equilibration of surface tension of certain solutions—the "surface-ageing" phenomenon—has been the subject of so many investigations that its essential validity cannot be questioned. It has recently been discussed in these *Reports* by Brown (1940). The anomaly is particularly

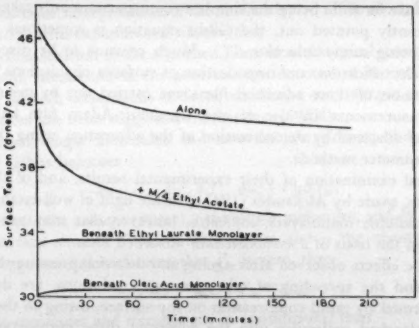


Figure 4. Effect of ethyl acetate and insoluble monolayers upon the "surface ageing" of sodium dodecyl sulphate (2×10^{-3} gm. mole) in N/100 HCl (from Alexander, 1941 a).

(Reproduced, by kind permission, from the 'Transactions of the Faraday Society'.)

evident with soap solutions, provided the concentration is below that at which micellar aggregation sets in; above this point equilibration is almost instantaneous.

Several explanations have been suggested. Doss (1935–39) and Nutting, Long and Harkins (1940) ascribe it to a type of activated adsorption, due to an electrostatic potential barrier formed by the adsorbed ions at the interface whereas McBain and Perry (1939) believe it to be due to formation of a "surface

pellicle", a bimolecular or polymolecular leaflet, with the hydrophilic groups not exposed, upon the surface.

Examination of the slow accumulation has recently been made by Alexander (1941a) at the air-water and oil-water interfaces. The effects of surface-active agents and of reducing the thickness of the oil phase in the latter case to molecular dimensions by "contaminating" the surface with insoluble, fluid monolayers were also investigated. The time factor was found to be reduced by addition of capillary-active substances such as ethyl acetate (see figure 4): at the oil-water interface or by covering the surface with "liquid expanded" monolayers such as oleic acid or ethyl laurate, it was completely eliminated (see figure 4).

Such experimental results are clearly inexplicable on the above theories, and Alexander (1941a) suggested, as a tentative hypothesis, that the rate determining step is the penetration and re-orientation, in the surface layer, of the hydrophobic portion of the molecule. Subsequent workers in this field (e.g. Hoffman, Boyd and Ralston, 1942a, 42b) have confirmed earlier experimental findings, but no alternative theories have been brought forward.

(ii) *The structure of the surfaces of solutions*

Considerable doubt has recently been cast, principally by McBain and his co-workers, upon the tacit assumption that the adsorbed layer upon the surfaces of solutions is monomolecular in thickness. Thus McBain, Ford and Mills (1940) write: "It is clear from all the phenomena referred to that the surfaces of solutions are far from being the simple structures commonly taken for granted. As was recently pointed out, the Gibbs equation is sometimes incompatible with their being monomolecular..... Much remains to be discovered about the depth, electrification, and organization of surfaces of solutions".

Examination of these adsorbed films was carried out by means of surface-tension measurements, by use of the Langmuir-Adam film balance, by the spreading of oils, and by determination of the adsorption using the microtome and interferometer methods.

A critical examination of their experimental results, and of those of other investigators, made by Alexander (1942c) in the light of well-established phenomena of insoluble monolayers, indicates, however, that they are equally well explained on the basis of a *monomolecular* adsorbed film.

Thus the effects observed after ageing and then compressing the surfaces of solutions, and the spreading of oils upon such solutions, are due to micro-crystals, formed by rapid compression on the surface, owing to the slow rate of escape of adsorbed molecules into the substrate, these micro-crystals then re-dissolving comparatively slowly.

In addition, the Gibbs adsorption was calculated on the basis of a monomolecular adsorbed film, and the result compared with the most recent experimental values from microtome and other measurements. With both the two compounds which have been studied in detail, namely hydrocinnamic and lauryl sulphonic acids, the calculated and experimental values agreed closely; for example, with the latter the calculated value was $5{\cdot}7 \times 10^{-10}$ g.mol./cm^2, the observed microtome values lying between 5·2 and $5{\cdot}7 \times 10^{-10}$ g.mol./cm^2.

The question of the apparent anomalies arising from the use of the Gibbs

equation are considered in the next section, where a simple explanation, also based upon a monomolecular adsorbed film, is put forward.

It would seem, therefore, that the old concept of a ***monomolecular*** adsorbed film can provide an adequate explanation for all the phenomena so far investigated.

(iii) *The Gibbs adsorption equation*

It is well known that anomalies arise when the Gibbs equation is applied to solutions of colloidal electrolytes, such as soaps and dyestuffs. (For recent reviews in these *Reports* see McBain and Mills (1938) and Brown (1940).) Thus the surface-tension/concentration curves of many such compounds show a horizontal portion in dilute solution, sometimes preceded by a minimum, at a surface tension which may be as low as 30 dynes/cm., presenting the paradox of a much reduced surface tension and yet a zero or negative surface excess of solute as calculated by applying Gibbs' fundamental equation in the usual manner.

Recent explanations have assumed either that the Gibbs equation is only a limiting law, and needs correcting terms to allow for surface orientation and the submerged double layer (McBain and Mills, 1938), or that the anomaly can be resolved by ascribing a particular molecular structure to the adsorbed layer (Long and Nutting, 1941). Both, however, are clearly fallacious, since Gibbs' treatment is thermodynamically rigorous and hence independent of any particular *molecular* interpretation of the surface layer. The second, in addition, is not in accord with experimental observation.

From a consideration of the adsorbed films of such typical "soaps" as lauryl sulphonic acid, mentioned in the previous section, and of the well-established aggregation of such compounds to form micelles, a very simple explanation of these anomalies has been suggested (Alexander, 1942 d).

When both single molecules and micelles are present the fundamental equation of Gibbs becomes

$$-d\gamma = \Gamma_w . d\mu_w + \Gamma_s . d\mu_s + \Gamma_M d\mu_M, \qquad \ldots\ldots(1)$$

where w denotes the solvent (water), s the single molecules, and M the micelles; γ denotes surface tension, Γ the surface excess and μ the chemical potential. Placing the geometrical surface so that Γ_w is zero, then

$$-d\gamma = \Gamma_s^{(w)} d\mu_s + \Gamma_M^{(w)} d\mu_M. \qquad \ldots\ldots(2)$$

Since single molecules and micelles are in equilibrium, then

$$\mu_s = \mu_M. \qquad \ldots\ldots(3)$$

Hence

$$-d\gamma = (\Gamma_s^{(w)} + \Gamma_M^{(w)}) d\mu_s, \qquad \ldots\ldots(4)$$

$$\Gamma_s^{(w)} + \Gamma_M^{(w)} = -\frac{1}{RT}\frac{d\gamma}{d\ln c_s f_s}. \qquad \ldots\ldots(5)$$

(For the dilute solutions considered c_s replaces N_s.)

The stoichiometric concentration

$$C = c_s + c_M, \qquad \ldots\ldots(6)$$

where c_s and c_M are expressed in equivalents.

If now $d\gamma/dC = 0$, as at the minimum of surface tension, then from equation (4)

$$\textit{either} \qquad \Gamma_s^{(w)} + \Gamma_M^{(w)} = 0 \qquad \ldots\ldots(7)$$

$$\textit{or} \qquad d\mu_s/dC = 0. \qquad \ldots\ldots(8)$$

Assuming $\Gamma_s^{(w)} + \Gamma_M^{(w)} \neq 0$, in agreement with experiment, then from equation (8), since $c_s f_s \neq 0$,

$$\frac{dc_s f_s}{dC} = 0. \qquad \ldots\ldots(9)$$

Hence $c_s f_s$ is independent of C, or approximately, the bulk concentration of single molecules is independent of the stoichiometric concentration.

Hence the fact that the surface-tension curve shows a minimum or horizontal portion does not *necessarily* imply, as hitherto assumed, that the total adsorption is zero, since it can equally well be explained by assuming that the concentration of single molecules becomes (approximately) independent of the stoichiometric concentration. This latter explanation is in excellent agreement with the conclusions obtained by applying the law of mass action to the single molecule-micelle equilibrium (Murray, 1935), and, being in agreement with experiment as well, resolves the apparent anomaly very simply.

§7. INSOLUBLE MONOLAYERS AT THE OIL-WATER INTERFACE

Despite the wide interest aroused by the behaviour of insoluble monolayers at the air-water interface, it was not until recently that one of the obvious extensions, namely, to the study of monolayers at the oil-water interface, was successfully attempted. The delay has been due more to the difficulties involved than to lack of interest, since the importance of an understanding of the phenomena at this interface, both for the biologist and from the viewpoint of emulsion theory and practice, needs no stressing (see, for example, Clayton, 1943).

The first recorded digression into this field is that of Askew and Danielli in 1936. Using a modification of the surface balance, they succeeded in measuring the Π-A relationship of a few compounds (egg-albumin, α-amino-palmitic acid, the C_{20} amide, and a methyl cellulose) at the bromobenzene-water interface. (See also Askew and Danielli, 1940.) Very great difficulties due to leaks and contamination were, however, encountered.

The problem was taken up by Alexander and Teorell (1939), who examined the Wilhelmy plate and ring methods, as well as the modified surface balance, for insoluble films at the benzene-water interface. Only the ring method was found to be satisfactory, the technique eventually evolved being outlined below.

A circular Pyrex glass dish was used, and after the water surface had been cleaned by suction, benzene was added and the interfacial tension (γ_0) determined from the maximum pull necessary to detach the ring. A known volume of solution of the substance under examination was then expelled into the interface by means of the usual "Agla" micrometer syringe. The new interfacial tension (γ) was determined as before, and hence Π, $= \gamma_0 - \gamma$, was found. The interfacial concentration was then increased by further additions of the spreading solution, not by decrease in the available area, as in the surface-balance method. Values of the area per molecule (A) are known from the area of the trough and the

amount of material spread, so that the Π-A relationship could be determined very simply. By this means errors due to leaks can be eliminated and an accuracy equal to that given by the surface balance for air-water monolayers may readily be obtained.

Accurate Π-A curves were thus determined for two proteins (gliadin and serum albumin), three phospho-lipoids (lecithin, lysolecithin and kephalin) (see also Alexander, Teorell and Åborg, 1939), and a typical detergent (sodium cetyl sulphate: see figure 5).

It was found in these cases that the ultimate packing of the molecules was precisely the same as at the air-water interface, indicating almost complete removal of the oil phase at high values of Π.

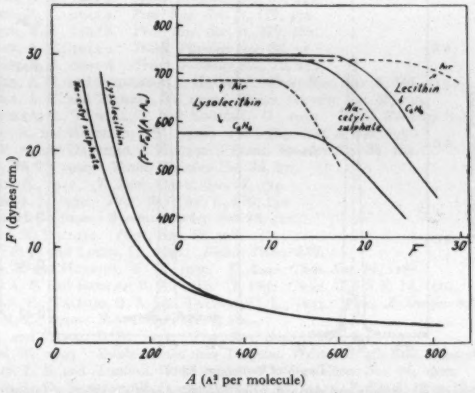


Figure 5. Force-area curves of lysolecithin and sodium cetyl sulphate at benzene-water interface.
Inset: $(\Pi-\Pi_0)(A-A_0)\sim\Pi$ at air-water and benzene-water interfaces
(from Alexander and Teorell, 1939).
(Note.—Surface pressures are here denoted by F and not by Π.)
(*Reproduced, by kind permission, from the 'Transactions of the Faraday Society'.*)

The Π-A curves fitted a hyperbolic equation $(\Pi-\Pi_0)(A-A_0)=C$ as shown in figure 5. In agreement with Langmuir's theory of expanded films (see p. 160), the Π_0 term is almost eliminated at the oil-water interface; with lysolecithin, for example, it falls from 12·8 dynes/cm. at the air-water to 0·05 at the benzene-water interface.

Determination of the change in phase-boundary potential (ΔV) produced by an insoluble monolayer is considerably more difficult at an oil-water than at an air-water interface, particularly if using pure hydrocarbon oils, owing to their high resistance. However, by using a very thin film of benzene, a large polonium electrode, and either a Lindemann or valve electrometer, Alexander and Teorell (1939) achieved a certain measure of success, and preliminary results

for lecithin have been published (Alexander, 1941 b). These results indicate freely mobile molecules with very small lateral cohesion, as suggested by the above equation of state, and a similar dipole orientation and value of ΔV (for a given molecular concentration) at both air-water and oil-water interfaces.

Measurements of the change in interfacial viscosity due to an insoluble monolayer are very few. The oscillation method (see p. 160) is suitable provided that a sufficiently thin film of the lighter phase is used, and a platinum needle replaces the usual oscillating vane. Changes in fluid films, to gels or solids, have thus been followed by Crisp (see Alexander, 1941 b, and figure 6). The rotating-disc method has been used by Pokhil (1939).

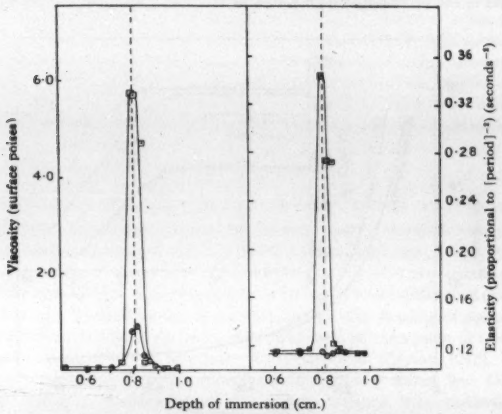


Figure 6. Interfacial viscosity and elasticity using oscillating needle. Effect of sulphate ions on long-chain amine hydrochloride film at bromobenzene-water interface.
- - - - - - - - - - indicates position of the interface.
⊙ before addition of sulphate. (Note low viscosity and absence of elasticity.)
⊟ after addition of sulphate. (Note high viscosity and marked elasticity.)

The above survey indicates that considerable advances in the field of oil-water insoluble monolayers should be possible now that suitable techniques are available. There are certain inherent advantages in the use of this interface—for example, proteins spread rapidly and completely, and compounds containing ionized groups, such as the soaps, are sufficiently stable for measurement. By such means it should be possible to extend and to clarify our understanding of interfacial phenomena in general, and of mobile interfaces in particular.

§ 8. CONCLUSION

If this *Report* has been able to convey some idea of the wide applicability of the results, and in particular of the general outlook, acquired from the study of insoluble monolayers, then it has served its purpose. It is, indeed, a fascinating

field, and I should like to take this opportunity of expressing my debt to Professor Rideal and to my numerous friends and colleagues, from this and other departments, at home and overseas, for the pleasure and benefit of stimulating discussions during the last eight years.

REFERENCES

ADAM, N. K., 1921. *Proc. Roy. Soc.* A, **99**, 336.

ADAM, N. K., 1941. *Physics and Chemistry of Surfaces*. 3rd Edit. (Oxford: The University Press).

AKAMATU, H. and SAMESHIMA, J., 1936. *Bull. Chem. Soc. Japan*, **11**, 791.

ALEXANDER, A. E., 1939. *J. Chem. Soc.* 777.

ALEXANDER, A. E., 1941 a. *Trans. Faraday Soc.* **37**, 15.

ALEXANDER, A. E., 1941 b. *Trans. Faraday Soc.* **37**, 117.

ALEXANDER, A. E., 1941 c. *Trans. Faraday Soc.* **37**, 426.

ALEXANDER, A. E., 1942 a. *Proc. Roy. Soc.* A, **179**, 470.

ALEXANDER, A. E., 1942 b. *Proc. Roy. Soc.* A, **179**, 486.

ALEXANDER, A. E., 1942 c. *Trans. Faraday Soc.* **38**, 54.

ALEXANDER, A. E., 1942 d. *Trans. Faraday Soc.* **38**, 248.

ALEXANDER, A. E. and SCHULMAN, J. H., 1937. *Proc. Roy. Soc.* A, **161**, 115.

ALEXANDER, A. E. and TEORELL, T., 1939. *Trans. Faraday Soc.* **35**, 727.

ALEXANDER, A. E., TEORELL, T. and ÅBORG, C. G., 1939. *Trans. Faraday Soc.* **35**, 1200.

ASKEW, F. A. and DANIELLI, J. F., 1936. *Proc. Roy. Soc.* A, **155**, 695.

ASKEW, F. A. and DANIELLI, J. F., 1940. *Trans. Faraday Soc.* **36**, 785.

ASTBURY, W. T., 1940. *Trans. Faraday Soc.* **36**, 871.

BERNSTEIN, S., 1940. *J. Amer. Chem. Soc.* **62**, 374.

BIKERMAN, J. J., 1939. *Proc. Roy. Soc.* A, **170**, 130.

BIKERMAN, J. J., 1940. *Trans. Faraday Soc.* **36**, 412.

BLODGETT, K. B., 1939. *Phys. Rev.* **55**, 391.

BOWDEN, F. P. and LEBEN, L., 1940. *Philos. Trans.* **239**, 1.

BOYD, G. E. and HARKINS, W. D., 1939. *J. Amer. Chem Soc.* **61**, 1188.

BRESLER, A. E. and BRESLER, S. E., 1940. *J. Phys. Chem. U.S.S.R.* **14**, 1604.

BRESLER, S. E., TALMUD, B. A. and TALMUD, D. L., 1933. *Phys. Z. Sowjet.* **4**, 864.

BROWN, R. C., 1940. *Rep. Prog. Phys.* **7**, 180.

CARY, A. and RIDEAL, E. K., 1925. *Proc. Roy. Soc.* A, **109**, 301.

CLAYTON, W., 1943. *Emulsions and their Technical Treatment*, 4th Edit. (Churchill).

COPELAND, L. E. and HARKINS, W. D., 1942. *J. Amer. Chem. Soc.* **64**, 1600.

COPELAND, L. E., HARKINS, W. D. and BOYD, G. E., 1942. *J. Chem. Phys.* **10**, 357.

COUMOULOS, G., 1943. *Proc. Roy. Soc.* A. In press.

CRISP, D. J., 1943. *Proc. Roy. Soc.* In press.

DERVICHIAN, D. G., 1939. *J. Chem. Phys.* **7**, 931.

DERVICHIAN, D. G. and JOLY, M., 1937. *C.R. Acad. Sci., Paris*, **204**, 1318.

DERVICHIAN, D. G. and JOLY, M., 1938 a. *J. Chem. Phys.* **6**, 226.

DERVICHIAN, D. G. and JOLY, M., 1938 b. *Nature, Lond.*, **141**, 975.

DERVICHIAN, D. G. and JOLY, M., 1938 c. *C.R. Acad. Sci., Paris*, **206**, 326.

DERVICHIAN, D. G. and JOLY, M., 1939 a. *C.R. Acad. Sci., Paris*, **208**, 1488.

DERVICHIAN, D. G. and JOLY, M., 1939 b. *J. Phys. Radium*, **10**, 375.

DOSS, K. S. G., 1935. *Curr. Sci.* **4**, 405.

DOSS, K. S. G., 1937. *Curr. Sci.* **5**, 645.

DOSS, K. S. G., 1939. *Kolloid Z.* **86**, 205.

FOURT, L. and HARKINS, W. D., 1938. *J. Phys. Chem.* **42**, 897.

GERMER, L. H. and STORKS, K. H., 1936. *Phys. Rev.* **50**, 676.

GERMER, L. H. and STORKS, K. H., 1938. *J. Chem. Phys.* **6**, 280.

GERMER, L. H. and STORKS, K. H., 1939. *Phys. Rev.* **55**, 648.

GIBBS, W., 1878. See *The Collected Works of J. Willard Gibbs* (London, New York: Longmans, Green and Co., 1931).

GLAZOV, N. I., 1938. *J. Phys. Chem. U.S.S.R.* **11**, 484.

GORANSEN, R. W. and ZISMAN, W. A., 1939. *J. Chem. Phys.* **7**, 492.

GREGG, S. J. and WIDDOWSON, E. E., 1939. *Nature, Lond.*, **144**, 666.

HARKINS, W. D., 1941. *Chem. Rev.* **29**, 385.
HARKINS, W. D. and BOYD, G. E., 1941. *J. Phys Chem.* **45**, 20.
HARKINS, W. D. and COPELAND, L. E., 1942. *J. Chem. Phys.* **10**, 272.
HARKINS, W. D. and KIRKWOOD, J. G., 1938 a. *J. Chem. Phys.* **6**, 53.
HARKINS, W. D. and KIRKWOOD, J. G., 1938 b. *Nature, Lond.*, **141**, 38.
HARKINS, W. D. and NUTTING, G. C., 1939. *J. Amer. Chem. Soc.* **61**, 1702.
HARKINS, W. D., YOUNG, T. F. and BOYD, G. E., 1940. *J. Chem. Phys.* **8**, 954.
HEYMANN, E., DOCKING, A. R., KERLEY, L. E. and MORTENSEN, K. N., 1940. *Nature, Lond.*, **146**, 265.
HEYMANN, E. and YOFFE, A., 1942. *Trans. Faraday Soc.* **38**, 408.
HOFFMAN, E. J., BOYD, G. E. and RALSTON, A. W., 1942 a. *J. Amer. Chem. Soc.* **64**, 498.
HOFFMAN, E. J., BOYD, G. E. and RALSTON, A. W., 1942 b. *J. Amer. Chem. Soc.* **64**, 2067.
HOLLEY, C. and BERNSTEIN, S., 1937. *Phys. Rev.* **52**, 525.
ISEMURA, T., 1940. *Bull. Chem. Soc. Japan*, **15**, 467.
JOLY, M., 1938. *J. Phys. Radium*, **9**, 345.
JOLY, M., 1939. *Kolloid Z.* **89**, 26.
KHEINMAN, A. S., 1940. *J. Phys. Chem. U.S.S.R.* **14**, 118.
KNOTT, G., SCHULMAN, J. H. and WELLS, A. F., 1940. *Proc. Roy. Soc.* A, **176**, 534.
LANGMUIR, I., 1933. *J. Chem. Phys.* **1**, 756.
LANGMUIR, I., 1934. *J. Franklin Inst.* **218**, 143.
LANGMUIR, I., 1938. *J. Amer. Chem. Soc.* **60**, 1190.
LANGMUIR, I., 1939. *Proc. Roy. Soc.* A, **170**, 1.
LANGMUIR, I. and SCHAEFER, V. J., 1937. *J. Amer. Chem. Soc.* **59**, 2400.
LONG, F. A. and NUTTING, G. C., 1941. *J. Amer. Chem. Soc.* **63**, 625.
MCBAIN, J. W., FORD, T. F. and MILLS, G. F., 1940. *J. Amer. Chem. Soc.* **62**, 1319.
MCBAIN, J. W. and MILLS, G. F., 1938. *Rep. Prog. Phys.* **5**, 30.
MCBAIN, J. W. and PERRY, L. H., 1939. *Ind. Engng. Chem.* **31**, 35.
MARSDEN, J. and RIDEAL, E. K., 1938. *J. Chem. Soc.* 1163.
MOORE, W. J. and EYRING, H., 1938. *J. Chem. Phys.* **6**, 391.
MOUQUIN, H. and RIDEAL, E. K., 1927. *Proc. Roy. Soc.* A, **114**, 690.
MURRAY, R. C., 1935. *Trans. Faraday Soc.* **31**, 206.
MYERS, R. J. and HARKINS, W. D., 1937. *Nature, Lond.*, **140**, 465.
NUTTING, G. C., LONG, F. A. and HARKINS, W. D., 1940. *J. Amer. Chem. Soc.* **62**, 1496.
POKHIL, P. F., 1939. *J. Phys. Chem. U.S.S.R.* **13**, 301.
PORTER, E. F. and WYMAN, J., 1937. *J. Amer. Chem. Soc.* **59**, 2746.
PORTER, E. F. and WYMAN, J., 1938. *J. Amer. Chem. Soc.* **60**, 2855.
RACE, H. H. and REYNOLDS, S. I., 1939. *J. Amer. Chem. Soc.* **61**, 1425.
SAMESHIMA, Z., AKAMATU, H. and ISEMURA, T., 1940. *Rev. Phys. Chem., Japan*, **14**, 55.
SCHULMAN, J. H., 1939. *Ann. Rep. Chem. Soc.* **36**, 94.
SCHULMAN, J. H. and RIDEAL, E. K., 1931. *Proc. Roy. Soc.* A, **130**, 268.
SEBBA, F. and BRISCOE, H. V., 1940 a. *J. Chem. Soc.* 106.
SEBBA, F. and BRISCOE, H. V., 1940 b. *J. Chem. Soc.* 114.
SEBBA, F. and BRISCOE, H. V., 1940 c. *J. Chem. Soc.* 128.
SEBBA, F. and RIDEAL, E. K., 1941. *Trans. Faraday Soc.* **37**, 273.
SKLYARENKO, S. I., BARANAEV, M. K. and MEZHUEVA, K. I., 1940. *J. Phys. Chem. U.S.S.R.* **14**, 839.
STÄLLBERG, S. and STENHAGEN, E., 1942. *J. Biol. Chem.* **143**, 171.
STENHAGEN, E., 1938. *Trans. Faraday Soc.* **34**, 1328.
STENHAGEN, E., 1940. *Trans. Faraday Soc.* **36**, 597.
SUTHERLAND, G. B. B. M. and TUTTE, W. T., 1939. *Nature, Lond.*, **144**, 707.
THIESSEN, P. A., BEISCHER, D. and VON GILLHAUSSEN, H., 1940. *Naturwissenschaften*, **28**, 265.
TRAPEZNIKOV, A. A., 1938. *Acta Physicochim. U.R.S.S.* **9**, 273.
TRAPEZNIKOV, A. A., 1939. *Acta Physicochim. U.R.S.S.* **10**, 65.
TRAPEZNIKOV, A. A., 1941. *Doklady Akad. Nank. U.R.S.S.* **30**, 319.
WIDDOWSON, E. E. and GREGG, S. J., 1939. *Nature, Lond.*, **143**, 760.
WILSON, R. E. and RIES, E. D., 1923. *Colloid-Symposium Monograph*, **1**, 145.
YAMINS, H. G. and ZISMAN, W. A., 1933. *J. Chem. Phys.* **1**, 656.
ZAHL, P. A., HASKINS, C. P., GALLAGHER, D. M. and BUCHWALD, C. E., 1939. *Trans. Faraday Soc.* **35**, 308.

FIELD EMISSION OF ELECTRONS

By R. O. JENKINS,*
The Research Laboratories of The General Electric Co., Ltd., England

§1. INTRODUCTION

THE extraction of electrons from cold metals by means of a strong electric field has been known for just over twenty years. During this time it has been referred to variously as auto-electronic emission, cold emission, or field emission. For convenience the last name will be used throughout this *Report*. Observations on the effect were published in 1922 by Lilienfeld, who caused a current of a few milliamperes to flow from a pointed electrode acting as a cathode in a high-vacuum x-ray tube. Since that time progress has been somewhat spasmodic compared with the continuous flow of research on the allied problem of thermionic emission. The probable explanation of this is that the emissions obtained were rather unreliable and, for various causes which have been found more recently, seldom gave very reproducible results. This fact limited their practical application, and the experiments published were thus much fewer than those on thermionics.

The work has, therefore, been connected much more closely to theory than has been the study of thermionic emission, where experimenters have often been satisfied with obtaining a useful practical result of their investigation without pursuing the theoretical aspects of their work very far.

This *Report* is an attempt to cover the ground that has been investigated since the early work by Lilienfeld, and may be divided approximately into three sections. The first one deals with the early investigations of 1921–30, during which time various experimenters sought to make accurate measurements of these field-emission currents, and relate them to some theory or express them empirically as a function of the field and temperature. This stage ended in a triumph for the then new wave-mechanics, which gave an explanation which fitted the known facts qualitatively, and subsequent measurements gave more complete agreement with the Fowler-Nordheim formula obtained from wave-mechanics.

In the intermediate period of 1930–36 relatively little work was done on the subject, possibly owing to the rival attractions of the quickly developing electron-optics, of electron-diffraction, and of the thermionic oxide cathode. All of these showed great promise as new and powerful techniques for use on other problems, and naturally attracted a large number of experimenters interested in harnessing the free electron for the service of science and the community.

The last period, which covers from 1936 to the present time, really starts with the work of Müller, who has made many notable additions to our knowledge. The advances in vacuum technique during the intervening years had

* The author takes pleasure in expressing to Mr. B. S. Gossling and other colleagues his thanks for their suggestions and interest in the production of this *Report*.

made it possible to obtain much more consistent emissions, and his results, together with his idea of using the experimental tube as a two-dimensional projection electron microscope, has stimulated experimenters in this country and in America to continue on these lines, with interesting and very beautiful results.

§2. EARLY EXPERIMENTS (1921–30) AND THE SCHOTTKY THEORY

(i) *Early experimental work*

The first important experimental work after the early observations of R. W. Wood (1897) was that of Lillenfeld (1922), who demonstrated that an x-ray tube with a pointed cold cathode passed a current of a few milliamperes when subjected to a strong electric field. A patent was taken out in 1919 and a fuller description of the tube was published in 1922. Lilienfeld also noted the enhancement of this field emission by contamination of the cathode by caesium. Other workers at the same time were investigating the effect, although many of their results were not published till much later. A patent of Gossling (1921) claimed the enhancement and stabilization of the field emission by coating the cathode with a layer of electropositive contamination, such as sodium.

This effect had also been noted previously by Dushman, Langmuir and Hull in 1914, although the results were not published until much later (Compton and Langmuir, 1930). This group continued their work, and Coolidge and Langmuir made more observations of the field emission in 1922, using as a cathode a broken piece of tungsten wire. The interest in this field-emission effect appears to have been somewhat diminished by the difficulty of obtaining reliable and repeatable results, and practical applications were similarly affected by its uncertainty.

(ii) *First theoretical explanation*

The first attempt to explain the emission of electrons from cold metals at high field strengths was made by Schottky (1923) in a long paper which attempted to correlate the various electron emissions, including the thermionic and field emission. The explanation followed the lines of the well-known Schottky effect in thermionic emission. This is the increase in thermionic emission with increasing field beyond the ordinary saturated emission given by the Richardson formula. The field emission was assumed to be caused by the same effect at considerably higher field strengths than those considered in the thermionic case, and the Schottky formula was extrapolated to a point where the emission at room temperature became significant. This indicated a field strength of the order of 10^8 volts/cm.

The derivation of the formula will be indicated here, as its assumptions and implications affected the subsequent experiments considerably. The work necessary to extract an electron from a metal was considered to be made up of two parts, the first due to atomic forces on the electron as it emerged from the lattice, and the second due to the image force as the electron continued its escape

from the surface. With no applied field the summit of the resulting potential barrier was at infinity. When the field was applied, the summit approached the surface to a point x_0 away from it (figure 1), and the height of the barrier decreased. The amount of this decrease was

$$\Delta\chi = -\frac{e^2}{4x_0} - eFx_0, \qquad \ldots\ldots(1)$$

where e is the electronic charge and F is the field, the first term being due to the decrease in work against the image force, and the second the work done on the electron by the field. The summit of the potential barrier is, of course, at the point where the force on the electron is zero, i.e. where the force due to the field equals that due to the image force. We have, therefore, that

$$\frac{e^2}{4x_0^2} = eF. \qquad \ldots\ldots(2)$$

Eliminating x_0.

$$\Delta\chi = -e^{3/2}F^{1/2}. \qquad \ldots\ldots(3)$$

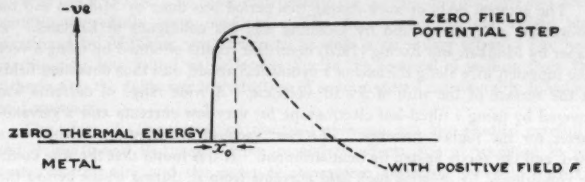


Figure 1. Potential barrier according to Schottky.

The new Richardson formula then becomes

$$i = AT^2 \exp\left(-\frac{\chi}{kT} + \frac{e^{3/2}F^{1/2}}{kT}\right). \qquad \ldots\ldots(4)$$

This formula has been verified by various workers on thermionic emission at relatively low fields. The assumption made in the derivation of this formula is the classical one used in the derivation of the Richardson formula, i.e. that only electrons with thermal energies greater than the height (χ) of the potential barrier can escape from the metal. The natural result of this is apparent in the formula, that the field emission should be dependent on temperature.

It may also be seen from (4) that a limiting case is reached when

$$\frac{\chi}{kT} = \frac{e^{3/2}F^{1/2}}{kT},$$

and the current then has a limiting value of AT^2. The rate of increase with temperature is then considerably less than the exponential rate of rise observed with the thermionic emission. Another implication is that at this limiting field strength the effective work function is zero, and no cooling due to the

evaporation of electrons should take place. By calculation from the known work function of tungsten, this limiting field should be of the order of 10^9 volts/cm. The field emission would only become appreciable at fields approaching this value, and this disagreed with the previous observations made by Lilienfeld and others, whose fields were of the order of 10^6 volts/cm. Schottky therefore suggested that the emission came from small geometrical irregularities on the cathode which could cause local fields considerably more intense than that experienced by the main body of the cathode. He calculated from idealized forms of excrescences that the field could be increased ten-fold by such a mechanism, and thus reconcile the theoretical field with that observed in practice.

This basis gave experimental workers some lines to follow, and the next phase of the subject was that of making more accurate measurements and attempting to correlate them to the Schottky theory.

(iii) *Later experimental work*

The greatest body of work during this period was done by Millikan and his co-workers in America and by Gossling and his colleagues in England. A paper by Millikan and Eyring (1926) described results obtained by using a very fine tungsten wire along the axis of a cylindrical anode, and thus obtaining fields at the surface of the wire of 5×10^6 volts/cm. A wide range of currents was covered by using a tilted-leaf electroscope for very low currents and a galvanometer for the higher currents. The best vacuum techniques available were used, and the anode heated by bombardment. It was found that the wire could be conditioned by drawing high field currents from it, during which period the current for a given field strength would decrease. After conditioning, repeatable curves could be plotted connecting current with field, providing the current did not exceed that used for conditioning. It was also observed that intense heating of the wire produced an increase in the field required to give appreciable current, but that this effect was not so permanent. The field currents were shown to have their origin in a few minute surface spots, as luminous spots would appear on the inside of the anode due to intense local bombardment. Their behaviour during conditioning could also be observed in this way. The authors concluded that the "current conditioning" of the wire was due to the smoothing of these submicroscopic irregularities by positive ion bombardment, while the heat conditioning was due to variation of contaminations on the wire. Finally, it was observed that the currents were independent of temperature over a range of 700° c., and this seemed to suggest that the electrons extracted were not "thermions" sharing the normal thermal energy of agitation of the atoms. It was also shown that the logarithm of the current did not bear a linear relationship with square root of the field as the Schottky equation predicted.

Subsequently Millikan and Lauritsen (1928) discovered that the results published in this previous paper could be plotted in the form of logarithm of the current against the reciprocal of the field, when an extremely good straight line, analogous to a Richardson thermionic plot, was obtained. Combining this result with the previous observation of slight increases in the current above 1100° K.,

they proposed the following formula to combine both field and thermionic emissions:

$$i = A(T + cF)^2 \exp\left(-\frac{b}{T + cF}\right), \qquad \ldots\ldots(5)$$

where F is again the field strength and A, b, c are constants.

This amounts to saying that the external field is equal to an increase in temperature of electrons in the metal. Millikan, Mackeown and Eyring (1928) in a subsequent paper extended their experimental work to the observation of field currents from points, and found again that very good straight-line plots were obtained between $\log i$ and $1/F$.

In the meantime, Gossling (1926) and his colleagues of the Research Staff of the G.E.C., London, had also been working on somewhat similar lines. Observations were made with cathodes of varying forms, such as a loop of wire, a brush of fine wires, or a single wire pointed by electro-chemical etching. Following the Schottky theory the measurements were plotted as $\log i$ against the square root of the potential difference between cathode and anode ($V^{1/2}$), but in no case were very good straight lines obtained, most plots being concave towards the $V^{1/2}$-axis. In these experiments the current-conditioning and heat-conditioning effects were observed. The fluorescence on the anode, also mentioned by Millikan, was shown to be only a secondary effect. The heat conditioning was found to be largely due to the removal of sodium from the cathode, and the deliberate introduction of sodium gave a considerably enhanced field emission.

The effect of temperature on the field emission was also studied carefully, and it was found that, apart from secondary effects due to variation in the contaminating film and the onset of thermionic emission, the field currents were independent of the temperature from 300–1500° K. Gossling gave a very full discussion of the theoretical implications of these results in connection with Schottky's theory, and came to the general conclusion that the experimental results did not give much confirmation of that theory, and that a better explanation might be found in terms of quantum rather than classical theory. It is interesting to note at this point that Gossling's figures, replotted later by Millikan, Mackeown and Eyring, agreed very well with Millikan and Eyring in giving the straight-line relation between $\log i$ and $1/F$.

De Bruyne (1928a) continued this work, using a spherical point made by fusing a thin filament, and showed that the emission was poisoned by the admission of small pressures of air, but only slightly by hydrogen at a similar pressure (10^{-3} mm. Hg). By evaporating barium on to the point the emission was considerably enhanced, only about one-third the previous applied potential being necessary to obtain a similar emission. In a later paper, De Bruyne (1928b) criticized the formula of Millikan and Lauritsen relating to the variation of field emission to temperature above 1000° K., and produced evidence to suggest that this variation was due to thermionic emission setting in. Millikan and Lauritsen (1929) replied that their previous evidence, together with a new experimental curve, was not at variance with De Bruyne's, which could equally well be interpreted by the Millikan-Lauritsen formula. Subsequent measurements by De Bruyne (1930) appeared to give much more conclusive evidence

that the field emission was really independent of temperature right up to the point where thermionic emission started. The increase in the emission noted by Millikan and his colleagues had apparently been due to the fact that the wire contained thoria, and after heat treatment would be partially activated with a surface layer of thorium, and could give appreciable thermionic emission (making allowance for the Schottky effect) at 1100° K.

Other workers had also been studying field emission during the same period with varying results. Del Rosario (1928) obtained negligible field emission from platinum filaments at fields of the order of 10^8 volts/cm., where other workers had found currents of the order of microamperes. If, however, a small pressure of gas were introduced, the currents then observed were a function of voltage on the tube, and not the actual field at the wire. It seems probable with our subsequent knowledge, that the surface of the wire must have been almost devoid of local irregularities which would have been necessary to give

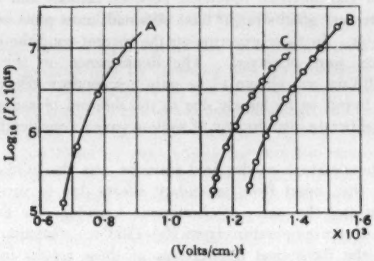


Figure 2. Log of field current plotted against square root of field (R. J. Piersol).

A. Before outgassing.
B. Outgassed 3 hrs. at 2700° K.
C. Outgassed 1 hr. at 2700° K.

appreciable field emission at these field strengths. Rother (1926) published a long paper describing work in which the high field strength had been obtained by extremely small clearances between the anode and cathode. These electrodes were supported by corrugated platinum membranes forming part of the vacuum envelope of the experimental tube. The tube was fixed in an interferometer by means of which the cathode-anode clearance could be accurately set and maintained at any desired value down to a few microns. The cathodes used were tantalum or iridium points or spheres. It was found that the field necessary to cause appreciable current to flow increased as the point was given a smoother surface finish by polishing, and also by heat treatment. A large number of experimental curves were published connecting current with applied voltage and cathode-anode clearance, and it was shown that for a "conditioned" point, fields of 8×10^{-6} volts/cm. were necessary to give appreciable currents. No attempt was made to correlate the observations with the Schottky theory apart from the fact that the field was of the order predicted by Schottky, and no attempt

was made to determine any empirical mathematical expression for the emission in terms of the field.

Piersol (1928), in America, using a tungsten filament on the axis of a cylindrical tube, obtained some neat and consistent results by employing very careful vacuum technique. The anode was a thin silver film on the inside of the glass envelope, and the tube was sealed off from the pump. It had a charcoal trap immersed in liquid air and an ionization gauge which indicated that the vacuum was of the order of 10^{-8} mm. Hg. The results obtained were in very good agreement with Millikan and Eyring, and curves obtained are shown in figure 2. The curves are for an untreated wire (A); the wire outgassed at 2700° K. for 3 hrs. (B), and outgassed for 1 hr. (C). Again it will be noted that the readings do not lie on a straight line as predicted by Schottky, and Piersol, in addition, showed that the emission was not changed by immersing the whole tube in liquid air, also in disagreement with Schottky theory.

At this stage of the *Report* it will have become very ovious that the classical Schottky theory, although well verified by De Bruyne (1928 c) and Pforte (1928) for thermionic emission, had failed to explain the field-emission phenomenon in two main respects, namely, the lack of variation of emission with temperature, and the relationship between emission and field strength. It was at this stage that the then comparatively new conceptions of mathematical physics came to the rescue in the form of quantum and wave mechanics.

§ 3. APPLICATION OF QUANTUM AND WAVE MECHANICS

Richardson (1928) and Houston (1928) attempted to explain these observed effects in terms of the new quantum theory, but the most generally accepted theoretical explanation is that due to Fowler and Nordheim, which will be outlined here. Nordheim (1928 a), in a paper chiefly concerned with thermionic emission, laid the basis for the explanation of field emission. He started with the Sommerfeld idea of a metal and assumed that the electrons in the metal had an energy distribution according to the Fermi-Dirac statistics, where the number of electrons with energies between E and $E+dE$ is

$$n(E)dE = 4\pi(2m)^{3/2}E^{1/2}/h^3\{1+\exp(E-\mu)/kT\}, \qquad \ldots\ldots(6)$$

where h is Planck's constant, m the mass of the electron, k is Boltzmann's constant in the metal, and μ the maximum kinetic energy of an electron at the absolute zero.

Nordheim then calculated the number of electrons per unit time arriving at the surface, which latter could be represented as a potential step (see figure 3). This involved an integration with respect to the electron energies normal to the surface, the problem reducing to one dimension only.

The number of electrons with normal energy W arriving per unit area of the surface in unit time was found to be

$$N(W) = \frac{4\pi mkT}{h^3} L\left(\frac{W-\mu}{kT}\right), \qquad \ldots\ldots(7)$$

where

$$L(\beta) = \int_0^\infty \frac{dy}{e^{\beta+y}+1}. \qquad \ldots\ldots(8)$$

In his original paper Nordheim considered only the thermionic case; the field-emission condition was worked out subsequently by Fowler and Nordheim (1928). They considered the potential barrier at the surface to be of the form shown in figure 3. This neglects the image force of the electron, which has the effect of rounding the top of the barrier. When the electrons to be considered are of comparatively low energy (which is true of the field electrons), this will not have much effect, but it does have a considerable effect on the emission coefficient for electrons with energies near the peak, which are, of course, those considered for thermionic emission under the Schottky high-field condition.

The electrons are then considered in terms of the Schrödinger wave equation and their behaviour found by solving the following equations :

$$\frac{d^2\psi}{dx^2} + K^2(W - C + Fx)\psi = 0, \quad x > 0, \qquad \ldots\ldots(9)$$

and

$$\frac{d^2\psi}{dx^2} + K^2 W\psi = 0, \quad x < 0, \qquad \ldots\ldots(10)$$

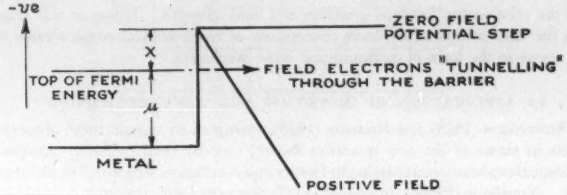


Figure 3. Potential barrier used by Fowler and Nordheim.

where F is the field, C is the "inner potential" and ψ is the probability function. The constant K^2 is given by

$$K^2 = 8\pi^2 m/h^2. \qquad \ldots\ldots(11)$$

These equations are subject to the conditions that ψ and $d\psi/dx$ are continuous at $x = 0$ and that for $x > 0$ (outside the metal surface) ψ represents a stream of electrons progressing to the right only (away from the metal surface).

Solving these wave equations, Fowler and Nordheim showed that the fraction of electrons with energy W which would escape through the potential hump was

$$D(W) = \frac{4\{W(C-W)\}^{1/2}}{C} e^{-4K(C-W)^{3/2}/3F}. \qquad \ldots\ldots(12)$$

As the number actually arriving at the surface from the inside of the metal is given by equation (7), the total current I escaping is given by

$$I = \frac{4\pi mekT}{h^3} \int_0^\infty D(W) L\left(\frac{W-\mu}{kT}\right) dW, \qquad \ldots\ldots(13)$$

where e is the electronic charge.

At ordinary temperatures $kTL\{(W-\mu)/kT\}$ will have an approximate value of $\mu - W$, when $W < \mu$ and is otherwise zero; the current thus becomes

$$I = \frac{16\pi me}{Ch^3}\int_0^{\mu} W^{1/2}(C-W)^{1/2}(\mu - W)e^{-\frac{4K(C-W)^{3/2}}{3F}}\,dW. \qquad \ldots\ldots(14)$$

and putting $C - \mu = \chi$, which is the normal thermionic work function, we have

$$I = \frac{e}{2\pi h}\frac{\mu^{1/2}}{(\chi+\mu)\chi^{1/2}}F^2 e^{-\frac{4K\chi^{3/2}}{3F}}$$

$$= 6{\cdot}2\times 10^{-6}\frac{\mu^{1/2}}{(\chi+\mu)\chi^{1/2}}F^2 e^{-\frac{6{\cdot}8\times 10^7\chi^{3/2}}{F}}, \qquad \ldots\ldots(15)$$

where I is in amps./cm^2, μ and χ are in volts and F is in volts/cm.

This fundamental equation is known as the Fowler-Nordheim equation for field-emission currents, and is the generally accepted relation with which to compare experimental observations. It will be noted that it bears a considerable resemblance to the Richardson thermionic equation, and is of the right form to agree with Millikan and Lauritsen's exponential plots of field emission, although it disagrees with their generalized formula, which also covered thermionic emission. As Fowler and Nordheim pointed out, this generalized formula was only verified for the limiting conditions of small field with high temperature and high fields with low temperature.

The physical picture involved in this equation is that when the field is very small, as in the case of thermionic emission, the potential barrier is a step as in figure 3, and to escape an electron must have enough thermal energy to raise it to the top of the step. When a very strong field is applied this step degenerates into a barrier which becomes progressively thinner as the field increases until there is a finite chance of the electron, with only the Fermi energy, escaping through the barrier, hence the term tunnel effect, which is often used when referring to the mechanism of escape of the field electron.

The Fowler-Nordheim equation indicated that the emission should become apparent at fields of the order of 10^7 volts/cm., which agreed quite well with observation, making allowance for the fact that the emission comes mainly from isolated surface irregularities where the actual field, as shown by Schottky, could be an order of magnitude greater than the measured field.

Nordheim (1928 b), in a subsequent paper, elaborated the calculation to include the effect of the image field. The corresponding wave equation for electrons escaping from the metal becomes

$$\frac{d^2\psi}{dx^2} + K^2\left(W - C + \frac{e^2}{4x} + Fx\right)\psi = 0. \qquad \ldots\ldots(16)$$

An approximate solution was obtained, and it was shown that, taking tungsten as a typical example in a field of 3×10^7 volts/cm., the power of the exponential in the Fowler-Nordheim formula (15) should be multiplied by 0·8, giving a corresponding reduction in field strength to produce the predicted field emission.

In this paper Nordheim also drew attention to the fact that in order to obtain a composite expression, covering both field and thermionic emission,

a rigorous solution of equation (16) would be necessary; he pointed out, however, that under most conditions the approximate calculations indicated that the two effects should be fairly well additive.

In a subsequent paper, Stern, Gossling and Fowler (1929) examined a set of experimental observations in the light of the Fowler-Nordheim formula. They assumed that, owing to the roughness of the surface, the actual field F was βF_m, where F_m was the measured field. From a plot of $\log_e I$ against $1/F_m$,

$$\frac{d\log_e I}{dF_m^{-1}} = -2F_m - \frac{6{\cdot}8\times10^7\chi^{3/2}}{\beta}, \qquad \ldots\ldots(17)$$

β was calculated, assuming χ.

Similarly, if A denote the emitting area,

$$\log_e A = 10{\cdot}0 + \log_e I - \log_e\frac{\mu^{1/2}}{(\chi+\mu)\chi^{1/2}} - 2\log_e(\beta F_m) - \frac{1}{F_m}\frac{d\log_e I}{dF_m^{-1}}. \qquad \ldots\ldots(18)$$

From plots of field emission from a fine electrolytically polished tungsten point β was found to be 2·68 and A $6{\cdot}3\times10^{-8}$ cm.^{2} By direct measurement the area was found to be $1{\cdot}2\times10^{-7}$ cm.^{2}, which was thus in quite good agreement.

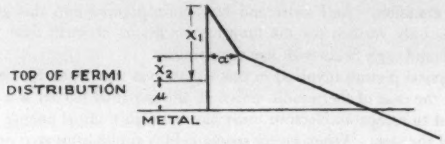


Figure 4. Potential barrier for contaminated surface.

It was found that when the point was contaminated by an electropositive material such as sodium, the slope of the plot altered, indicating a reduction in χ, but it also caused a marked reduction of the apparent emitting area A, which seemed inexplicable on any simple theory. Further theoretical calculations were therefore made assuming that, as the new surface was a composite one, consisting of a layer of sodium atoms on tungsten, the potential barrier would be of the form shown in figure 4, where a would be of the order of atomic dimensions. It was found that if χ_1 was the work function of the clean metal and χ_2 that of the composite layer, equations (17) and (18) became

$$\frac{d\log_e I}{dF_m^{-1}} = -2F_m - \frac{6{\cdot}8\times10^7}{\beta}\chi_2^{3/2}, \qquad \ldots\ldots(19)$$

$$\log_e A = 10{\cdot}0 + \log_e I - \frac{\log_e\mu^{1/2}}{(\chi_1+\mu)\chi_1^{1/2}} - 2\log_e(\beta F_m) - \frac{1}{F_m}\frac{d\log_e I}{dF_m^{-1}} - \log_e\frac{\chi_1}{\chi_2} + \frac{4Ka[\chi_1+\chi_2+(\chi_1\chi_2)^{1/2}]}{\chi_1^{1/2}+\chi_2^{1/2}}. \qquad \ldots\ldots(20)$$

The last two terms of (20) give the natural logarithm of the ratio of the apparent emitting areas of the clean and contaminated surface. Gossling took field-emission plots of a tungsten point in the clean state, and after it had been

partially and fully activated with sodium. From the change in slope of the plots the work function for the sodium on tungsten surface was found to be 1·4 volts. From the ratio of the apparent emitting areas for the clean and sodium activated point, a value of $2{\cdot}2 \times 10^{-8}$ cm. was deduced for the thickness of the film, which agrees quite well with the known size of the sodium atom.

§4. EXPERIMENTAL WORK IN THE YEARS 1930–36

During this period, which was so fruitful in other directions of electronic research, very little of importance was discovered in field emission, and the number of papers published was relatively few. Two papers of interest were those of Chambers and Ahearn, who continued work on much the same lines as previous experimenters. Chambers (1934) sought to discover the discrepancies between the work of Del Rosario (1928) and the other groups of previous workers. Del Rosario had found that he was unable to obtain emission at fields comparable with other workers, and Chambers repeated his experiments, using the working parts of Del Rosario's tube. In the light of results on electrical breakdowns in high vacuum devices obtained by Bennett (1931), he discovered that the heat treatment and vacuum conditions used by Del Rosario were such that there could have been no small projections or ruptures on the wire cathode to give the local high field strengths from emitting spots. What little emission Del Rosario had obtained came from the filament supports. These anomalous results were thus correlated with those of other observers, and added evidence to the general validity of the Fowler-Nordheim formula.

Ahearn in a previous paper in 1933 had obtained further evidence confirming De Bruyne's (1930) conclusion that the field currents were independent of temperature up to the point where the current was completely masked by the thermionic emission. In his later paper (Ahearn, 1936) he extended his previous work to composite surfaces by using thoriated tungsten as the cathode, and again found that the emission was independent of the temperature up to the point where thermionic emission was very great compared with field emission.

A surprising observation was that the field currents were independent of the degree of thoriation as deduced from thermionic emission from the whole filament. Ahearn suggested that this did not necessarily disagree with the Fowler-Nordheim formula, which required that the emission should vary rapidly with the work function, because the small projections forming the emission centres might not be activated to the same extent as the main surface. Further work on electrical breakdown in high vacuum indicated that such breakdowns were intimately connected with the field emission and might be caused by the ohmic heating of the emitting spots by field current which in turn would reduce the tensile strength of the material at that point to below the electrostatic pull due to the high field. Gossling (1932; see also Bell, Davies and Gossling, 1938), however, denies any necessary connection between field emission and breakdown. About the same time Wehnelt and Schilling (1935) obtained an electron microscope image of field emission from a molybdenum plate, and showed that the emission was coming from sub-microscopic points which were not permanent but would disappear with temperature or increase in activity with increases in the applied field.

A series of papers was published in this period on the field emission from mercury surfaces, arising from problems involved in starting and maintaining arc discharges in mercury arc rectifiers. Beams (1933) found, with a steel sphere 0·25 mm. away from the surface of a pool of mercury, that the field required to start an arc discharge increased from 6×10^5 volts/cm. to $1{\cdot}8 \times 10^6$ volts/cm. by repeated redistillation of the mercury. To avoid distortion of the surface, the potential was applied as a microsecond pulse, and from other data it was found that 1 amp./cm.^{2} field emission should be necessary to start an arc in this short time. To account for even this very small current at the observed field strength Beams postulated that there must be surface irregularities similar to those causing field emission from solid metals. Quarles (1935) continued this work and introduced a platinum filament into the tube. He used this as a cathode and measured the contact potential between it and the mercury, thus deducing variations in the work function of the mercury surface. The field necessary to start the arc discharge was plotted against the change in work function. These results suggested that, within the accuracy of the experiment, the field emission was varying more rapidly than was predicted by the Fowler-Nordheim formula. Moore (1936) further extended these observations to higher field strengths with very similar results.

§5. EXPERIMENTAL WORK IN THE YEARS 1936–43

(i) *Electrical measurements*

In 1936 Müller, working in the Siemens Laboratories at Berlin, published a long paper in which he examined the field emission from clean and activated tungsten points and wires in the light of the Fowler-Nordheim theory. This paper (Müller, 1936 a) marked the beginning of a new interest in field emission, and the results he obtained were also summarized in two papers appearing in other publications (Müller, 1936 b, 36 c). Müller studied field emission along two main lines, firstly to find the variation of the emission with work function of the surface, and secondly to confirm that the escaping electrons came from the top of the Fermi energy distribution in the metal and thus well below the top of the potential barrier at the surface. Two experiments confirmed the latter fundamental assumption of the Fowler-Nordheim theory. If the electrons which escaped had appreciable thermal energy in the metal, we should expect a cooling of the point, due to the loss of this thermal energy, similar to that observed when a high thermionic current is drawn from a hot cathode such as tungsten. Müller calculated the cooling that might be expected to take place at the tip of a small tungsten point when a field current was taken from it. He showed that if the field electrons had thermal energies in the metal of only 0·1 electron-volts, they would cause a 10° drop in temperature at the point when a field current of 5 milli-amps. was drawn from it. This should have been observable with an optical pyrometer when the initial temperature of the point was in the region of 1000–1100° K. In the actual experiment no cooling of the point was observed at all. The other method of showing that the field electrons came from an energy level well below the peak of the potential barrier at the surface was to take a distribution curve of the energies of electrons after they

had left the metal. This was achieved by placing the point cathode at the centre of a hemispherical anode of molybdenum mesh. A concentric hemispherical nickel cup surrounded the outside of the anode and was held at a potential within a few volts of that of the cathode. By varying this potential and observing the current it collected, it was found that the cup needed to be 5 volts positive to the point before any electrons reached it. The current saturated at +10 volts. This indicated that the energy of the field electrons in the tungsten must have corresponded to the top of the Fermi level μ and not to the inner potential $(\chi+\mu)$. The 5 volts necessary to collect the electrons was thus approximately equal to the work function of the collector. as explained later in this *Report*.

For investigating the variation of field emission with work function of the surface, Müller used a tube with a fine tungsten point facing an anode plate with a 0·7-mm. hole. Barium could be evaporated on to the point through this hole by heating with high-frequency eddy currents a small nickel cup containing the barium. Immediately adjacent to the point was a small tungsten filament used for measuring the work function, and also as a cathode to bombard the point and remove surface impurities. When the tube had been outgassed and evacuated, a field-emission plot of $\log_e I$ against $1/F$ was taken, and the tungsten filament then used as a thermionic cathode and its work function determined. It was assumed that the point work function was the same. Barium was then evaporated on to the point and filament, and the field-current plot from the point and the Richardson plot of the filament taken. This was done for various amounts of barium up to a monatomic layer.

The natural logarithm of the slope of each field-current line ($\log_e I$ against $1/F$) was plotted against $\log_e \chi$ obtained from the Richardson line and a straight line of slope 3 was obtained. This indicated that χ^3 appeared in the exponential of the Fowler-Nordheim formula, which disagreed with their prediction of $\chi^{3/2}$. $\mathrm{Log}_e A$ was also found to have a linear relation with χ. The experiment was repeated using a thin filament as the field and thermionic cathode, and contaminating it with more than a monatomic layer of barium.

A similar relationship of field emission with work function was observed. Experiments using magnesium and caesium were made, but did not give very conclusive evidence in this point.

In a subsequent paper Müller (1937 a) studied the effect of adsorption of thorium on a point cathode. In this experiment he used an etched tungsten point of 1 micron radius, and had a piece of thorium sintered on to a tungsten filament from which to evaporate thorium on to the cathode. He found that the slope of field-emission plots again gave χ^3 as the exponential in the field-emission formula, assuming the known values of work function for thoriated and clean tungsten. He was able to show from a series of such plots the variation of work function as the thorium was evaporated from the cathode. He started with a work function of 3·3 volts for a more than monatomic layer of thorium, decreasing to 2·7 volts for the monatomic layer, and increasing again up to 4·8 volts for clean tungsten as the thorium was evaporated by raising the point to increasingly high temperatures. The removal of the film by ion bombardment was studied by introducing small quantities of argon, hydrogen and nitrogen, and the poisoning of the thorium layer by oxygen was also observed. It was

again found in these experiments that $\log_e A$ had an approximately linear relationship with χ.

At this stage of the *Report* it is convenient to consider the work of Henderson and his collaborators before dealing with the use of the field-emission projection microscope. Henderson and Badgley (1931), and later Henderson and Dahlstrom (1939), measured the distribution of energies in the field electrons. Their apparatus consisted of a thin, straight, filamentary cathode on the axis of a perforated cylindrical anode. A concentric cylinder outside the first acted as a collector of the electrons. In figure 5 a schematic diagram of the potentials of the electrons is shown, and it may be seen that the electron leaving the cathode tunnels through the thin potential barrier. As the field at the collector is small, the barrier is very thick, and the electron thus must have enough energy to climb over the barrier before it can reach the collector. This can only be done by making the collector positive to the cathode by an amount equal to its work function. Electrons from levels lower than μ can only reach the collector if the anode is proportionally more positive. The field emissions obtained in this experiment were very stable, and the vacuum technique carefully controlled. The experimental curves indicated that virtually none of the field electrons

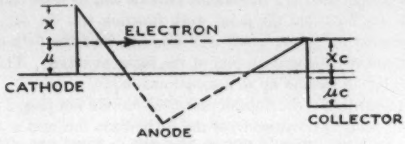


Figure 5. Potentials in energy-distribution experiment.

had energies in the cathode above the Fermi level, and that the bulk had energies between 5 and 10 volts, which agreed well with the Fowler-Nordheim theory. Theoretical curves were also plotted and were shown to have a similar form to the experimental ones.

Later work of Abbott and Henderson (1939) covered accurately a large range of field-emission current and field-strength measurements, and they were thus able to show the slight curvature in the $\log_e I$ against $1/F$ plots caused by the presence of the F^3 term in the Fowler-Nordheim formula. They showed, however, that the term was more correctly F^4, their experimental results being expressed as

$$I = AF^4 \exp(-b/F).$$

They pointed out that this might not be in disagreement with the theoretical formula, which assumed a constant field over the whole emitting area. This was never obtained in practice, and the observed power of F would depend on the microscopic shape of the point. A later paper by Fleming and Henderson (1940) described the measurement of the cooling effect on a cathode caused by the escape of electrons as field and thermionic currents. This was achieved by making a tungsten-tantalum thermocouple a part of the cathode. It was shown that the loss in heat due to field currents was negligible, confirming Müller's

result, and agreeing with theory, while a value of 4·46 volts was obtained for the work function of tungsten as calculated from the loss in heat due to the "evaporation" of electrons.

(ii) *Field-emission projection microscope*

In 1936 Johnson and Shockley (1936) constructed a very simple one-dimensional electron microscope for studying thermionic emission. It consisted of a thin filament mounted axially in the centre of an anode consisting of a wire helix. This fitted inside a cylindrical glass tube, the inner walls of which were coated with fluorescent material. The emission from the filament thus gave an image on the walls of the tube which was a projection of the emission surface, true size along the axis of the tube, and magnified around the tube in a ratio of tube diameter to wire diameter. Müller (1937b) adopted this scheme for studying the field emission by taking a fine etched tungsten point surrounded by a cylindrical anode. The tip of the point faced a fluorescent screen on the end wall of the glass vacuum envelope at a distance of about 10 cm. from the point. The field emission, coming radially from the approximately spherical end of the point, then struck the fluorescent screen, and gave a projected reproduction or "image" of the emission from the surface of the point. This image was enlarged in the ratio of the distance between point and the screen to the radius of point. The point was made by etching in sodium nitrite and was of the order of one micron radius. The magnification was thus of the order of 10^4, while the potential necessary to obtain an image was approximately 10,000 volts, and depended on the size of the point and the state of its surface.

Müller found that with an untreated tungsten point the field emission came from sharp lines. When the point was strongly heated, by passing current through a tungsten loop to which it was attached, the emission became more uniform over the surface, but there were always certain parts which did not emit, and appeared as circular dark patches on the screen. He found that these corresponded to certain directions in the single crystal forming the point, the largest dark patch corresponding to a [110] plane, four smaller ones grouped around it to [211] planes, and a small dark spot to the [100] plane. A somewhat similar picture was obtained for molybdenum, which has, of course, a similar body-centred cubic crystal lattice (see plate 2). On admitting oxygen to the tube, preferential poisoning took place in the region surrounding the [100] plane, so that when the applied field was raised by 80% to obtain an image, the whole of this region remained dark. By heating the point at progressively higher temperatures the removal of the oxygen layer was followed till, after heating at 2200° K., the original pattern was obtained at the original field strength. Two pictures from nickel and copper points were also shown, each exhibiting the [111] plane as a non-emitting area with brightly emitting parts surrounding it, but Müller found it extremely difficult to obtain any useful results from these points, due to their instability. This he attributed to the relatively low melting point of the metals, which precluded any thorough outgassing treatment. He indicated that this type of tube should be very useful for observing the behaviour of absorbed layers, and included a picture obtained from a partially thoriated tungsten point which showed that the thorium layer was confined to a relatively

small percentage of the surface area and that the individual thorium atoms or groups of atoms could be observed.

Müller (1938) continued his work in a second paper, making observations on the mobility of tungsten atoms on the surface of the point. He observed that the emission from the freshly etched point came from sharp lines which were the edges left between the main crystal planes etched on the surface of the single crystal point. On raising the temperature, the emission could be observed to smooth out, and actual movement of the tungsten atoms repositioning themselves could be observed, starting at a temperature of 1100° K. The movement of the atoms was ordered by the underlying lattice structure into definite lines of flow. After extremely strong heating, the pattern corresponded to a uniform emission over the whole surface except for the non-emitting main planes [110], [211] and [100].

Müller suggested that these spots might be due either to intrinsically high work function or to depressions in the otherwise approximately spherical surface. Another phenomenon observed was that the pattern obtained when the point was heated with the field applied had a number of more highly emitting parts which could not be due to contamination, but must have been caused either by small projections formed on the surface, or by some kind of orientation of the tungsten atoms on the surface giving a lower work function. This point will be discussed later.

To investigate the cause of the non-emitting spots on the smooth points, Müller fired barium from a source facing one side of the point, and thus condensed a barium layer on that side of the point. From the fact that the non-emitting portions of the surface could be made to emit under these conditions Müller deduced that there could not be cavities sufficiently deep to suppress the emission when the point was clean. At the same time he made a number of interesting observations on the mobility of the barium atoms on the surface. He showed that when a more than monatomic layer was deposited, migration of barium atoms over the tungsten started at 690° K., while the thick layer formed itself into crystallites which broke up again at a somewhat higher temperature. By heating for a few seconds at 1500° K. only a very sparse covering of barium was left, and the individual adsorbed ions were visible as brightly emitting spots. From this work it was clear that the technique had great advantages over the thermionic method for studying migration phenomena, since the condition of the surface could be observed in the cold state between any desired heat treatments.

Benjamin and Jenkins, who had previously (1938) been studying the migration of barium and thorium on tungsten and molybdenum, using the Johnson and Shockley microscope, adopted the Müller apparatus in a slightly modified form (figure 6) for continuing their work to clear up anomalies between their results and those of Becker (1933, 40). While agreeing with Becker on the migration of thorium, they had not been able previously to confirm the migration of barium, but observed evaporation before appreciable migration took place. Their first paper (1940) was concerned with the field emission from clean tungsten, molybdenum and nickel. Their vacuum technique differed from that of previous workers in that they worked with sealed-off tubes, which were of a hard glass

that could be baked to 700° C., and that they used a very efficient source of pure barium to provide a getter layer on the bulb well away from the fluorescent screen. The fluorescent screen was free from binder, and these precautions should have given a degree of vacuum better than that of other workers. Their observations on tungsten agreed well with those of Müller in showing mobility above 1170° K. They obtained more marked modifications to the pattern when the point was heated in the presence of the field. They obtained the most striking evidence from molybdenum points, however, where the mobility commenced at 770° K., and with the field applied higher mobilities could be obtained before thermionic emission commenced. A set of patterns is shown in plates 2–4. Plate 1 is an original etched point showing strong emission from the edges between crystal planes, and plate 2 another point in the smooth state after flashing without the field. Plate 3 is the latter after heating to 1520° K. with the field applied, and the potential necessary to obtain the picture had by that time fallen to about 80% of its original value. After heating to 1770° K. with the field applied, the pattern had undergone a very marked alteration

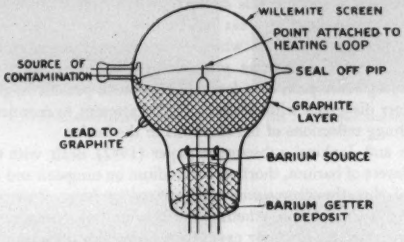


Figure 6. Field-emission projection electron microscope.

(plate 4), and the potential had then dropped to 60% of its original value. On heating without the field the point reverted to the state shown in plate 2. Benjamin and Jenkins, with the evidence of a field plot for the points corresponding to plates 2 and 4, considered that actual distortion of the shape of the point was caused by the presence of the electrostatic field, whose force on the surface was becoming comparable with the strength of the metal, and with surface tension. When mobility was present, the surface shape altered to a new configuration corresponding to a state of minimum energy with the field present. On "freezing" the point with the field applied this shape was retained. Confirming this conclusion, pictures have since been taken by Dr. G. D. Preston with an electron microscope. These are shown in plates 5 and 6,* the points corresponding to plates 2 and 4. It will be seen that the point in the case of plate 2 was extremely smooth, while the distortion of the shape which would cause the two highly emitting spots on plate 4 can be seen on the picture of the point in plate 6. The irregular-shaped dark masses attached to the point in

* The electron micrographs (plates 5 and 6) were taken in the Metallurgy Department of the National Physical Laboratory, and it is desired to thank the Director of the Laboratory for his permission to reproduce them (*Crown Copyright is reserved*).

plate 6 are pieces of foreign matter picked up during removal of the point from the tube.

A series of pictures from a nickel point using the field-emission tube were also obtained. These exhibited the building-up phenomenon to an even more marked degree, and mobility became apparent at 370° K. One of the patterns is shown in plate 7, and it will be observed that the symmetry is different from that of molybdenum or tungsten, the non-emitting planes being in this case the [110], [100], [111] and [113]. It was found quite easy to cause very strong building-up on the [111] plane by heating up to 1120° K., and if too much field were applied the point was disrupted. This appears to agree well with Ahearn's suggestion concerning the mechanism of electrical breakdown *in vacuo*.

A discussion concerning the cause of the non-emitting parts of the point led to no definite conclusions. In agreement with Müller, Benjamin and Jenkins thought that small flats or pits in the surface were not adequate to explain the negligible emission, but a firmly adsorbed oxygen film at these points, which could not be removed, might give the necessary increase in work function. An observation which supported this suggestion was that when the vacuum was not perfect, these non-emitting areas became larger, but returned to their normal size again on heating to 650° K., when the adsorbed gases were removed from the surface. Another possibility put forward was that the Brillouin zones might be playing some part, or a lattice interference pattern might be formed, but the authors disagreed with Müller's (1939) attempt to correlate the pattern with direct Bragg reflections of the electrons in the lattice.

Benjamin and Jenkins' subsequent paper (1942) dealt with the behaviour of adsorbed layers of barium, thorium and sodium on tungsten and molybdenum. They showed that the disposition of the thorium layer on a fully activated tungsten point was the same whether the thorium was evaporated on to the point from an external source, or came from reduction of thorium oxide in the tungsten. Migration of the thorium atoms over the tungsten surface started at 870° K. and evaporation at 2100° K. The pattern from the fully activated surface is shown in plate 8, where the field was only 60% of that necessary for the clean point. The rings of atoms are around the [110] plane, the bright patch in the centre of the triangle formed by these three rings is the [111] plane, while the eight small bright spots grouped round the [100] plane correspond to the [421] planes. On evaporation the thorium was most persistent on the [111] plane. The thorium layer on molybdenum behaved in a somewhat similar manner, but on raising the temperature to 1700° K. the thorium disappeared except on the [421] planes, either due to inward diffusion or evaporation (plates 13 and 14).

Their observations on the behaviour of barium films on tungsten agreed very well with those of Müller in showing migration at 700° K. and recrystallization in thick films (plates 15 and 16). When a monatomic layer was present it preferred the [211] planes, and evaporated first around the [110] plane (plates 9 and 10). They did, however, find that, if a certain amount of oxygen were also evolved with the barium, the migration did not take place, but the barium appeared to form pools which slowly evaporated as the temperature rose. When barium was evaporated on to a smooth molybdenum point, much the same behaviour as

Plate 1. Untreated molybdenum point.

Plate 2. Flashed and smooth molybdenum point.

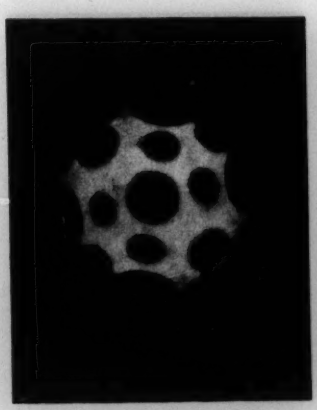


Plate 3. Molybdenum point "built up" by heating at 1520° K. with field applied.

Plate 4. Molybdenum point fully "built up" by heating at 1770° K. with field applied.

To face page 194

Plate 5. Electron micrograph of smooth molybdenum point (×15,000).

Plate 6. Electron micrograph of fully "built-up" molybdenum point (×15,000)

Plate 7. Nickel point in partially "built-up" state.

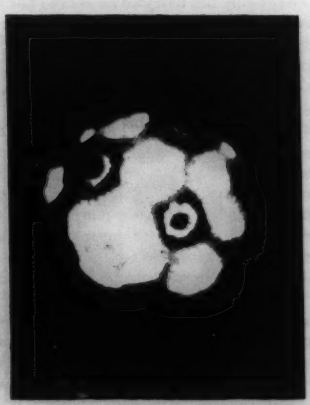


Plate 8. Tungsten point fully activated with thorium.

Plate 9. Barium layer on tungsten spreading at 770° K.

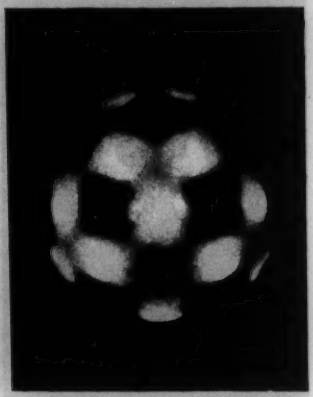


Plate 10. Barium layer commencing to evaporate from tungsten around the [110] planes at 1000° K. while covering the [211] planes.

Plate 11. Initial barium deposit on "built-up" molybdenum point.

Plate 12. Barium layer covering "built-up" molybdenum point after heating at 700° K.

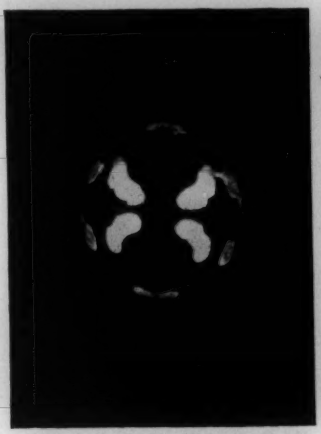


Plate 13. Thorium on molybdenum after heating to 1200° K.

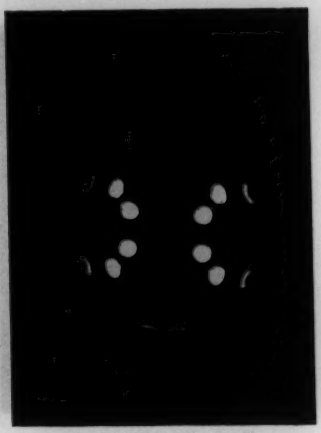


Plate 14. Thorium on molybdenum after heating to 1700° K. Thorium only on {421} planes.

Plate 15. Heavy deposit of barium on tungsten.

Plate 16. Heavy deposit of barium on tungsten heated to 700° K. Migration of a monatomic layer and formation of crystallites has taken place.

on tungsten was observed, the fully activated point giving a similar pattern to the clean point, but at one-third of the voltage. When barium was evaporated on to a "built-up" point the typical "built-up" pattern persisted, but again at the reduced voltage (plates 11 and 12). The authors considered this as strong evidence in favour of the supposition that the changes in pattern during the "building-up" process were due to mechanical distortion of the surface causing local high fields. The evidence of the movement of the barium, and also of the possibility of its inhibition by oxygen, cleared up the discrepancies between the previous work of Becker and of Benjamin and Jenkins.

The behaviour of sodium on molybdenum and tungsten was very similar. If the layer was less than monatomic, migration occurred at 500° K., but if more than a monatomic layer was deposited, migration took place at room temperature. The sodium appeared to avoid the [110] zone, but covered the [211] zone in a manner similar to barium on tungsten. Evaporation commenced at 600° K. and was complete at 950° K. The mobility of the thick film at room temperature confirmed the work of Bosworth (1935), who observed the movement photo-electrically.

The pattern obtained from a sodium-activated "built-up" molybdenum point was again similar to that from the clean "built-up" point, but at about 40% of the field. An obvious feature was that the strongest emission on the activated smooth point was from the [211] plane, which appeared as a non-emitting part on the "built-up" point, again suggesting that the actual shape of the point had changed during "build up".

At approximately the same time as the last paper Daniel (1942) described similar experiments on the "build-up" phenomenon on a tungsten point, where the effect was much less marked than with molybdenum. He attributed it to migration of surface impurity. From his description it appears that his vacuum conditions were not as good as those of Müller or of Benjamin and Jenkins; some of the effects he observed were correctly interpreted. His observations on thorium-activated tungsten points and the pictures published were very similar to those of Benjamin and Jenkins. Daniel had also published a previous paper (1941) on observations of field-emitting spots on tungsten filaments in a Johnson and Shockley type of one-dimensional electron microscope.

In an interesting paper Haefer (1940) described the use of the ordinary electron microscope for finding the size and shape of the point, and obtained pictures similar to plate 5. By substituting this observed area and radius into the Fowler-Nordheim formula he obtained a theoretical value for the field emission from clean tungsten which agreed very closely with the observed value. From this he deduced that the 3/2 power of the work function was confirmed. He made experiments with barium, potassium and caesium layers which also appeared to agree with the formula, and suggested that Müller's different results had been caused by crystallite formation.

§6. FUTURE TRENDS

It is apparent that physics and physical chemistry have, in the projection electron microscope, a tool of far-reaching importance. There are still, in the

domain of field emission, a number of important fundamental points to be cleared up, the first being the interpretation of the non-emitting parts on a clean smooth point. The evidence of the ordinary electron microscope, and also experiments with adsorbed films, appear to confirm that these are not due to pits or other mechanical distortions of the surface, although such do occur when the point is heated with the field applied. The projection patterns obtained by Martin (1939) from a polished single-crystal tungsten sphere used as a thermionic source show extraordinary similarities to the field-emission patterns, but with less contrast between the emitting and non-emitting parts. This would suggest some form of variation of work function over the surface of the metal, although the magnitude appears to be too small to account for the field-emission patterns. The other fundamental point is a really complete verification of the Fowler-Nordheim formula. The latest vacuum techniques, such as Benjamin and Jenkins used, and the actual measurement of the size of the point by means of the ordinary electron microscope, together with the emission being made visible, so that extraneous effects can be eliminated, should enable the formula to be verified, and allowances made for variation of emission in different directions.

REFERENCES

Abbott, F. R. and Henderson, J. E., 1939. *Phys. Rev.* **56,** 113.
Ahearn, A. J., 1936. *Phys. Rev.* **50,** 238.
Beams, J. W., 1933. *Phys. Rev.* **44,** 804.
Becker, J. A. and Brattain, W. H., 1933. *Phys. Rev.* **43,** 428.
Becker, J. A. and Moore, G. E., 1940. *Phil. Mag.* **29,** 129.
Bell, J., Davies, J. W. and Gossling, B. S., 1938. *Proc. Inst. Elec. Engrs.* **83,** 176.
Benjamin, M. and Jenkins, R. O., 1938. *Phil. Mag.* **26,** 1049.
Benjamin, M. and Jenkins, R. O., 1940. *Proc. Roy. Soc.* A, **176,** 262.
Benjamin, M. and Jenkins, R. O., 1942. *Proc. Roy. Soc.* A, **180,** 225.
Bennett, W. H., 1931. *Phys. Rev.* **37,** 582.
Bosworth, R. C. L., 1935. *Proc. Roy. Soc.* A, **150,** 58.
Chambers, C. C., 1934. *J. Franklin Inst.* **218,** 463.
Compton, K. T. and Langmuir, I., 1930. *Rev. Mod. Phys.* **2,** no. 2.
Daniel, J. H., 1941. *J. Appl. Phys.* **12,** 645.
Daniel, J. H., 1942. *Phys. Rev.* **61,** 657.
De Bruyne, N. A., 1928 a. *Phil. Mag.* **5,** 574.
De Bruyne, N. A., 1928 b. *Proc. Camb. Phil. Soc.* **24,** 518.
De Bruyne, N. A., 1928 c. *Proc. Roy. Soc.* A, **120,** 428.
De Bruyne, N. A., 1930. *Phys. Rev.* **35,** 172-176.
Del Rosario, C., 1928. *J. Franklin Inst.* **103,** 205.
Fleming, G. M. and Henderson, J. E., 1940. *Phys. Rev.* **58,** 887.
Fowler, R. H. and Nordheim, L. W., 1928. *Proc. Roy. Soc.* A, **119,** 173.
Gossling, B. S., 1921. *British Patent No.* 195, 440.
Gossling, B. S., 1926. *Phil. Mag.* **1,** 609.
Gossling, B. S., 1932. *Proc. Inst. Elec. Engrs.* **71,** 460.
Haefer, R. H., 1940. *Z. Phys.* **116,** 604.
Henderson, J. E. and Badgley, R. E., 1931. *Phys. Rev.* **38,** 540.
Henderson, J. E. and Dahlstrom, R. K., 1939. *Phys. Rev.* **55,** 473.
Houston, W. V., 1928. *Z. Phys.* **47,** 33.
Johnson, R. P. and Shockley, W., 1936. *Phys. Rev.* **49,** 436.
Lilienfeld, J. E., 1922. *Phys. Z.* **23,** 506.
Martin, S. T., 1939. *Phys. Rev.* **56,** 947.
Millikan, R. A. and Eyring, C. F., 1926. *Phys. Rev.* **27,** 51.
Millikan, R. A. and Lauritsen, C. C., 1928. *Proc. Nat. Acad. Sci., Wash.*, **14,** 1, 45.
Millikan, R. A., and Lauritsen, C. C. 1929. *Phys. Rev.* **33,** 598.

MILLIKAN, R. A., EYRING, C. F. and MACKEOWN, S. S., 1928. *Phys. Rev.* **31**, 900.
MOORE, D. H., 1936. *Phys. Rev.* **50**, 344.
MÜLLER, E. W., 1936 a. *Z. Phys.* **102**, 734.
MÜLLER, E. W., 1936 b. *Phys. Z.* **37**, 838.
MÜLLER, E. W., 1936 c. *Z. tech. Phys.* **17**, 412.
MÜLLER, E. W., 1937 a. *Z. Phys.* **106**, 132.
MÜLLER, E. W., 1937 b. *Z. Phys.* **106**, 541.
MÜLLER, E. W., 1938. *Z. Phys.* **108**, 668.
MÜLLER, E. W., 1939. *Naturwissenschaften*, **49**, 820.
NORDHEIM, L. W., 1928 a. *Z. Phys.* **46**, 833.
NORDHEIM, L. W., 1928 b. *Proc. Roy. Soc.* A, **121**, 626.
PFORTE, W. S., 1928. *Z. Phys.* **49**, 46.
PIERSOL, R. J., 1928. *Phys. Rev.* **31**, 441.
QUARLES, L. R., 1935. *Phys. Rev.* **48**, 260.
RICHARDSON, O. W., 1928. *Proc. Roy. Soc.* A, **117**, 719.
ROTHER, F., 1926. *Ann. Phys., Lpz.*, **81**, 317.
SCHOTTKY, W., 1923. *Z. Phys.* **14**, 80.
STERN, T. E., GOSSLING, B. S. and FOWLER, R. H., 1929. *Proc. Roy. Soc.* A. **124**, 699.
WEHNELT, A. and SCHILLING, W., 1935. *Z. Phys.* **98**, 286.
WOOD, R. W., 1897. *Phys. Rev.* **5**, 1.

PHYSICS AND THE SEARCH FOR OIL

By J. McG. BRUCKSHAW,
Imperial College, London

§1. INTRODUCTION

THE subject of geophysical prospecting is not confined only to the search for oil, but has many important applications in base-metal prospecting, water location, and so forth; however, since the publication of the last *Report* on this subject (Lancaster-Jones, 1936), the major developments and successes of this type of prospecting have undoubtedly occurred in connection with oil. Accordingly, it seemed desirable to review the progress in this branch in reasonable detail rather than to survey the whole field of geophysical prospecting in a rather inadequate manner, due to limitations of space. Of the various methods of prospecting, those most suitable for the search for oil are the seismic method, the gravitational method, and, to a lesser degree, the magnetic method. The development in the first method is almost entirely due to the introduction of the reflection technique, and in the second method to the appearance of modern gravity meters. These instruments, suitable for field use, are often capable of detecting changes in gravity of 0·1 milligal or less, a sensitivity which compares very favourably with that required by the British Association, which, in 1888, offered a prize for a portable field instrument to measure changes of 10 milligals. In addition to these two major developments, the remanent magnetism of rocks has been found useful for the orientation of drill cores, an application of physics to one of the important problems associated with drilling.

§2. THE SEISMIC METHODS

Probably the most outstanding development in prospecting made in recent years is the reflection method of seismic prospecting, and it has been employed successfully in many areas in the U.S.A., South America and elsewhere for the location of geological structures suitable for the accumulation of oil. Prior to its introduction, the gravitational method, and later the seismic refraction method, had been employed, and these had yielded important results; both methods, however, suffer from grave disadvantages. The former method cannot be operated in areas of rough topography, or in regions where flat-lying or gently-dipping strata occur. The first of these disadvantages arises from the difficulty of estimating accurately the gravitational effects of the surface irregularities at the various observation stations, particularly if the Eötvös torsion balance, measuring the horizontal gradient of gravity, is employed; the second defect is due to the very small effects produced by this type of geological structure. The refraction seismic method overcomes both of these objections, in fact in regions of gentle dips it functions most satisfactorily, but in its early form it was incapable of yielding fine detail, and many small structures could not be examined

by the method. Finally, before the refraction method can be applied, one essential condition must be satisfied, namely, the underlying medium must transmit mechanical disturbances with a velocity greater than the overlying rocks. The reflection method uses detectors close to the source of the mechanical disturbance, so that to a first approximation the depth determinations may be regarded as applying to a point on the surface. Hence, by a suitably close network of observation points, quite small details of structure may be revealed. In addition, any discontinuity of velocity or density in the rock formation will lead to a reflection of the mechanical disturbance, so that the severe limitation of the velocity condition is removed. Against this, no information is obtained concerning the nature of the medium below the reflecting horizon, and for the depth measurements the velocity must be determined by subsidiary experiments. In refraction work the procedure used for the depth observation also yields the values of the velocities involved, and hence some indirect information concerning the nature of the rocks.

(i) *The reflection technique*

Although reflection surveys were carried out as early as 1926 (Rosaire and Adler, 1934; Weatherby, 1940), there was a considerable interval before reliable details were published. Included amongst the earliest descriptions is a paper by Goldstone (1934), and many of the basic ideas discussed by him are still employed. There were two main difficulties to be overcome before the method could be employed: to accentuate the record of the reflected pulse, so that it could be picked out against the background of other arrivals and the general ground unrest, and then to develop a criterion for identifying it as a reflected pulse. A buried explosion, usually used as the source of energy, sets up a number of disturbances, a longitudinal wave, a transverse wave and a surface wave, the Rayleigh wave, these being in order of decreasing velocities. Of necessity, the reflected pulse can never be the first arrival, and for observation stations close to the spot point the fastest longitudinal reflection from a deep-seated horizon will arrive later than the direct surface wave, which carries a large fraction of the energy of the explosion and which, in addition, persists for a long time. Compared with the intensity of the surface wave, which decreases inversely with the distance, that of the reflected pulse is small, since only a few per cent of the incident energy is reflected (Knott, 1899; Muskat and Meres, 1940 a, 40 b), and the intensity decreases inversely as the square of the path length. Fortunately there is also an absorption of energy by the medium (Howell, Kean and Thompson, 1940; Born, 1941), an absorption which varies with frequency, with the result that the disturbance, which is initiated in the form of a sharp pulse of short duration, arrives at the detectors in the form of a damped train of waves. For the surface disturbance, the " ground roll ", the frequency of the waves is usually less than 30 per second, but for the reflected pulse it is considerably higher, the major part of the energy being carried within the frequency range of 40–80 per second (Weatherby, 1940). This fact is utilized in the separation of these two components. Electrical seismographs, or geophones, are always employed in reflection work, and their output, of the same frequency as the ground motion, is transmitted through

high-pass or band-pass filters, so designed that there is ample discrimination between the wanted frequencies, which are transmitted practically unaltered, and the low frequencies, which are suppressed. The selectivity can be further increased by the use of suitable amplifiers before passing the output to the recording instruments, usually multi-string galvanometers or oscillographs. This technique leads to a record on which a number of arrivals can be identified, but there is nothing to indicate which of them corresponds to a reflected pulse.

To identify the reflections, the direction from which they arrive is utilized. A number of geophones are placed on a straight line, usually passing through the shot point, and are joined through the separate filters and amplifiers to a central recording station, where all records are taken on the same film. Any surface wave, or refracted wave from shallow boundaries, will excite the instruments in succession, the time interval between adjacent instruments depending on the separation and the velocity of propagation of the rock through which the wave has passed. If the instruments are at equal intervals and produce traces in sequence across the film, the separation between the traces being equal, the arrivals will fall on a straight line whose slope gives a measure of the apparent surface velocity. The important reflected pulses are those from deep-seated boundaries, and these arrive at a steep angle from below, so that the time interval between the excitation of neighbouring geophones is small, corresponding to a high apparent surface velocity. If the electrical characteristics of the recording channels are similar, and the geophones have identical response characteristics, practically identical traces will be obtained from all channels. Thus the criterion for a reflected pulse is a set of identical records having practically the same arrival time. Since the wave front is curved, and, from dipping reflecting horizons in particular, the reflected ray is rarely vertical, there is a time-difference in the arrivals varying up to about 0·04 second between the excitation of the first and last instrument. This time-difference is referred to as the "step-out" of the reflected pulse.

(ii) *Interpretation of reflection surveys*

From the time of travel of the reflected pulse and its "step-out" the depth and dip of the reflecting horizon can be computed, if the velocity at all points along its path be known. The method of utilizing the results, however, depends on the geological conditions obtaining in the area surveyed. The two principal methods are: (*a*) the method of correlation, or depth shooting, and (*b*) dip shooting. The conditions necessary for the former are (i) a good reflecting horizon producing an outstanding reflection which can be readily identified on the records at all observation stations; (ii) the reflecting horizon must persist throughout the area of the survey; and finally (iii) the horizon must conform with the general geological structure of the region, so that its depth variations are significant. These conditions are found to occur in Oklahoma, where the contacts between limestones and shales form a suitable horizon for this type of prospecting. The depths to the various contacts are measured at a suitable number of stations and the structure is revealed by simple correlation. The structure is represented either by a contour map of the reflecting horizon or by sections. Figure 1 shows an example of the latter, and in it two reflecting

boundaries reveal the existence of faults (McDermott, 1932). More frequently, the geological conditions are not so simple as this, and a sequence of young sedimentary beds do not contain any outstanding reflecting boundaries, but rather a large number of horizons of very similar reflecting properties. In addition, due to the overlapping and lenticular nature of the beds, any one reflecting interface only persists for a limited horizontal extent, and the beds thin out and are replaced by others. In this case, if the measured depths below the stations are plotted, an apparently random distribution of points is obtained, and there is no indication of how they should be correlated to reveal the reflecting boundaries and the structure. If, however, at each depth-point a short line is drawn at the measured angle of dip, the tendency lines of the structure stand out (figure 2) and the presence of any anticline shown up, and, in some cases,

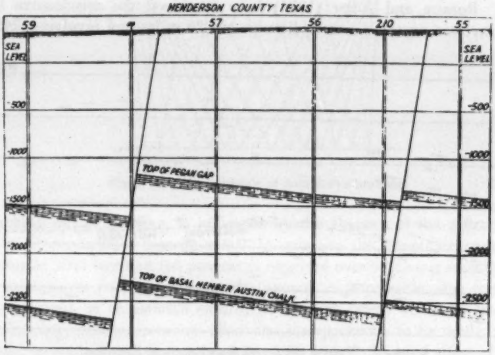


Figure 1. Example of correlation shooting.
(*Reproduced, by kind permission, from Amer. Assoc. Petr. Geol., Bull.*, 1932.)

the correlation between certain reflection horizons becomes possible. This method has been employed in the Texas-Louisiana-Gulf-Coast region.

The earliest successes were obtained in Oklahoma, where the correlation method was employed. Initially, it was anticipated that the same methods could be applied elsewhere, with the result that the early efforts in the Gulf-Coast region were not very successful. The use of the step-out for the purpose of dip determinations must be credited to T. I. Harker (Rosaire and Adler, 1934), who noticed that abnormal values were obtained, and correctly attributed the low values to the condition of shooting up the dip and the large values to shooting down the slope.

The method of dip shooting, in its simplest form, consists of carrying out a series of observations along a number of traverses, usually radiating from a point. For each traverse, the form of the geology may be examined by a section of the type shown in figure 2. A second method of depicting the results is available.

Starting at some convenient depth at a point, usually at the intersection of a number of traverses, a phantom horizon (i.e. an artificial persistent horizon which can be used to reveal the major structural features) is constructed, using the average values of the measured dip at each observation station, and from a large number of intersecting traverses the horizon can be contoured. This system is sufficient so long as no closed traverses are made, but when the survey includes such traverses, the computed depth of the phantom horizon will show a closing error, and in certain cases the misclosure may have some geological significance. Part of the error will arise from observational errors, both random and systematic; but since the dip observations are made at discrete intervals, such features as faults or terraces between the stations will not be taken into account. Further, if the throw or step changes between the two intersections of the feature with the closed traverse, part of the misclosure will be due to the feature. Rosaire and Adler (1934) have examined the misclosures in 151 closed traverses covering a total distance of 786 miles and involving 2458 dip

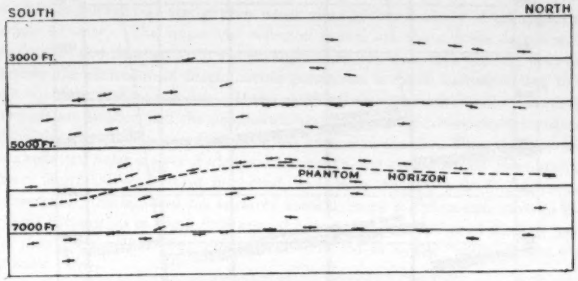


Figure 2. Typical dip profile from Southern Louisiana.
(*After Goldstone*, 1934.)

determinations. They conclude that the probable error for the closure is 87 feet in traverses averaging 27,500 feet in length. Of the errors arising from the method, the random errors in the dip tend to cancel, while errors due to incorrect velocities can be detected and average out by placing shot points at equal distances on opposite sides of the geophones. Within limits, neither the spacing of the stations nor the lengths of the traverse appear to be significant in controlling the misclosure.

In the correlation method, the existence and throw of faults can be inferred from a discontinuity in the measured depths arising from observations on each side of the fault, and by additional observations in the appropriate interval the position of the fault can be obtained accurately. In the dip method, however, the presence of a fault is not revealed directly and its position is obscure. By a modification of the normal procedure, described independently by Pirson (1937) and Salvatori (1937), the details of the thinning and disappearance of beds and the abrupt depth change at a fault can often be examined even though

no persistent key horizon exists. The principle is illustrated in figure 3, in which G_1 G_2 .... are geophones placed at equal intervals in a straight line and PQ the reflecting horizon. For a shot fired at G_1 and recorded on geophones G_2 to G_9, the reflections occur over the section PR of the interface, while for a shot at G_9, recorded on G_1 to G_8, the section RQ is operative. The two shots, accordingly, use the whole length of the reflecting surface, and, in addition, the records can be correlated, since the path G_1RG_9 is common to both. In this way a continuous profile of the bed is obtained, and the end of the bed, due

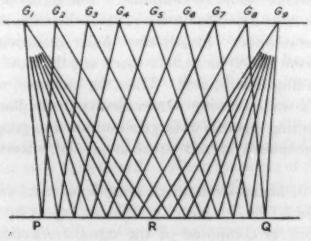


Figure 3. The principle of continuous profiling.

to pinching out or faulting, is indicated by the absence of the reflection on a number of consecutive instruments. In practice the geophone spread is of the order of 1000 feet and the process is repeated over adjoining sections. At a fault, continuous correlation over the record is impossible, due to the discontinuity, and, as an isolated record, the interpretation would be impossible. The adjoining records, however, allow the interpretation to be made, although it should be appreciated that with a large fault hade * the total effect of the fault may occupy a large number of records, and the system becomes more complex than the simple principle suggests.

(iii) *Multiple recording and variable compounding of records*

The method of obtaining records with the technique described above frequently leads to records of poor quality, on which it is difficult to identify the same type of wave form from one trace to another. This, as has been shown, may be due to the presence of a fault, which may limit the number of instruments receiving a reflection from a particular boundary, but this is by no means true in all cases of poor correlation between traces. Considerable improvement in quality, and of the ease with which the records may be interpreted, can be obtained by multiple recording. In this a number, n, of geophones, of similar response characteristics, are placed on a straight line at equal intervals, and are all joined in series with one recording channel. This gives the recording system a directional selectivity, for obviously a wave arriving vertically from below will excite all geophones at the same instant, and the resultant output will be n times

* The angle between the fault plane and the vertical

the output of a single geophone. On the other hand, if the wavelength of a surface arrival is equal to the spread of the geophones in one recording channel, there is a regular interval in the phases of the geophone outputs, leading to destructive interference between them. Use has been made of this fact to form an additional method for the suppression of the unwanted ground roll and at the same time to increase the signal level, the spread in any one channel being selected to correspond approximately to the wavelength of the "ground roll" Naturally the ground roll is not eliminated entirely thereby, since it consists of a damped train of waves, rather than a continuous sine wave, for which the destructive interference would be complete. The actual direction characteristics of the geophone system depends on the ratio of the channel spread to the wavelength of the disturbance. Steady-state polar sensitivity curves have been calculated by Johnson (1939) for various cases, and the most favourable conditions for multiple recording determined. Thus, for example, maximum discrimination in favour of a wave arriving perpendicular to the line of the geophones is obtained by connecting together as large a number n of geophones as convenient, and making the geophone spread equal to $(n-1)^2/n$ times the effective wavelength.

The question of signal-strength to noise-level ratio for multiple recording has been approached from an entirely different aspect by Klipsch (1936). He regarded the record as composed of the signal (reflection), the regular noise and the irregular noise, the last arising from the action of wind, microseisms and other general ground unrest, and bearing no relation to the shot. It is obvious that the irregular noise can be treated statistically, since it is random in nature, but by making suitable assumptions the regular noise, i.e. the unwanted background produced by the shot, can also be treated in a similar manner. Each impulse produced by the disturbance can be resolved into a spectrum of various frequencies, each frequency consisting of a damped train of waves, the damping of the train and its attenuation during transmission through the rocks depending on the frequency. The various arrivals have travelled along different paths with various velocities, since many types of waves are involved, and at any one point a large number of arrivals occur. The resultant motion is treated as being random in nature, this involving the assumption that, after firing the shot and during transmission, there are a number of independent causes occurring at random to modify the characteristics of the pulse. Indirectly, there is some experimental justification for this since two consecutive shots fired under identical conditions do not produce identical records. Thus, if a certain ground motion takes place at a point, an almost similar motion will take place at a neighbouring point some time later, the main difference between the two being expressible as a phase difference of the various frequency components. Due to the modifying action of the independent causes, the phase difference for any component is not fixed, and all phase differences within certain limits are equally probable. From this it can be shown that, using n geophones suitably disposed at the surface, instead of a single instrument, the average gain in the signal-noise ratio is given by $2(n/\pi)^{\frac{1}{2}}$

Even with the introduction of multiple recording, areas have been examined in which no reliable reflections have been obtained, leaving gaps in the data

obtained. Once again, such areas are frequently associated with faulting, but other geological structures may have similar effects. Riebar (1937) has suggested that the lack of identifiable reflected pulses can be explained by the interference between two practically simultaneous arrivals having different directions. Such conditions are possible over the axis of a syncline, in the vicinity of an unconformity, or near a fault. The separation of these simultaneous arrivals appears possible with the geosonograph developed by Riebar (1936). In this, one geophone only is connected to each recording channel, and the record, instead of consisting of a trace of varying displacement from a zero line, is a variable-density track of the same nature as a sound track on a film. A number of parallel records, each from a separate geophone, are obtained on the same film. The record tracks are at equal intervals on the film, corresponding to equal distances between the geophones, and they can be compounded by passing a beam of light through a suitably designed slit placed across the film, the resultant intensity of the transmitted light being continuously recorded, by a photo-electric cell and recording galvanometer, as the film is traversed past the slit. The angle made by the slit with the film is variable, and a series of compounded charts for different slit angles is made from each field record. When the angle of the slit corresponds to the "step-out" of an arrival, the separate records of the reflection are added in phase, all other arrivals with a different step-out going to form the general background level. Accordingly, as the slit angle approaches and passes through the appropriate step-out, the charts (recorded side by side for ease of inspection) show a peak gradually rising out of the background to a maximum, and then subsiding back again. Since different arrival directions correspond to different step-outs, these can be separated if they are sufficiently different. By this method of variable compounding, the directional selectivity is the same for all directions of arrival, and the same result would be obtained if, in multiple recording, it were possible to place the geophones in any one channel on a line perpendicular to the arrival direction. Mott-Smith (1937) has criticized the method, mainly on the grounds that variable compounding introduces more background than multiple recording, and also that the step-out cannot be measured accurately. Johnson (Mott-Smith, 1937), however, has shown that the accuracy of the step-out measurement compares favourably with that obtained in the normal method, and the increase in background is not very serious.

(iv) *Basis of depth and dip calculations*

The geometry of the reflection method has been worked out by various investigators. In the simplest case, if T is the travel time for a reflected pulse from a horizontal boundary, V the constant velocity along its path, X the distance from the explosion point to the geophone, then the depth p of the reflecting horizon is given by

$$V^2T^2=4p^2+X^2, \qquad \ldots\ldots(1)$$

and, since frequently the depth is large compared with X, the approximation

$$p=VT/2 \qquad \ldots\ldots(1\,a)$$

is often used. The more important problem is the determination of the dip, and for this the time "step-out" over the geophone spread is utilized. Again,

if p is the perpendicular distance from the reflecting horizon dipping at an angle θ, then

$$V^2T^2 = 4p^2 + X^2 - 2pX\sin\theta, \qquad \ldots\ldots(2)$$

in which the other symbols have the same significance as before, and X is positive in the direction of decreasing depths, i.e. up the slope. For a suitable geophone spread equations can be obtained using the values of T and X for the extreme instruments, and hence both θ and p can be determined. The exact solution, however, is not suitable for rapid and easy computation, and various approximate formulae have been suggested, frequently employing special geophone arrangements to give simplification. Thus, for two geophones at equal distances X on opposite sides of the shot point,

$$4p^2 = V^2(T_1^2 + T_2^2) - 2X^2$$

and

$$\sin\theta = V^2T\Delta T/2pX, \qquad \ldots\ldots(3)$$

where T is the mean arrival time and ΔT the time difference.

Gutenberg (1936) suggested the formulae

$$p = VT_0/2$$

and

$$\sin\theta = X/VT_0 - VT\Delta T/T_0\Delta X, \qquad \ldots\ldots(4)$$

in which the various symbols have the same meaning as above, except that X is now the ***mean*** distance of the geophones from the shot point, while ΔX is their maximum separation (the geophone spread) and T_0 is the arrival time of the reflected wave at the shot point, usually obtained by graphical extrapolation. This same formula, derived in a different manner, was put forward by Pentz (1937), who also examined its validity. For slopes less than 0·25 the maximum error is 0·015 if the reflecting boundary is deeper than 2500 feet and the distance of the farthest geophone less than 2500 feet. One important fact emerges from these formulae, for the dip calculation always involves ΔT, the step-out on the record. Since this is of the order of 0·01 second only, and at the best can be estimated to the nearest 0·001 second, the dip can only be measured to the nearest degree or so.

At very great depths, and with complex geology, it is frequently difficult to determine the velocity accurately, and both depths and dips become indeterminate. Fortunately, useful information is often made available from a knowledge of the direction of dip. With geophone distances $x_1, x_2, x_3, \ldots$ and arrival times $t_1, t_2, t_3, \ldots$, the direction of dip can be ascertained (Specht, 1940) from the sign of the expression

$$\frac{t_2}{x_2}\cdot\left(\frac{t_3-t_1}{x_3-x_1}\right) - \frac{t_4}{x_4}\cdot\left(\frac{t_5-t_3}{x_5-x_3}\right). \qquad \ldots\ldots(5)$$

A positive value of this expression indicates that the direction of x is down the slope, and for a negative value it is up the slope. For a zero value of the expression the reflecting boundary is horizontal. It is claimed that this criterion is sufficient, as well as necessary, to give the direction of dip. Here again it will be noticed that small time differences are involved and, for gently dipping strata, some confusion may arise unless the accuracy of the expression is carefull scrutinized in each case.

In general, the velocity is not constant between the surface and the reflecting boundary, and the above formulae must be modified to take this important factor into account. It was noted by Ewing and Leet (1932a, 32b) that certain time-distance curves obtained in refraction prospecting were better represented by smooth curves instead of the usual series of straight lines normally used for their interpretation. The latter system, used because the necessary formulae were available for the interpretation, is based on the assumption of a succession of homogeneous layers, each with a constant velocity of propagation, and the approximate nature of this picture was appreciated much earlier (Barton, 1929). In the case where the ground is composed of a series of such beds, each of uniform thickness and having a constant velocity of propagation, the equation to any ray can be written in the form (Slotnick, 1936a)

$$x = \sum_{1}^{n-1} h_i \tan \alpha_i + h \tan \alpha_n \qquad \ldots\ldots(6)$$

and

$$t = \sum_{1}^{n-1} h_i / v_i \cos \alpha_i + h / v_n \cos \alpha_n,$$

where h_i is the thickness of layer i in which the velocity is v_i, α_i is the angle made between the wave-path and the vertical in this layer, and h the distance of the point of the ray considered from the upper boundary of layer n. Since the refraction at each boundary obeys Snell's law, each ray may be defined by a parameter p, given by

$$p = \frac{\sin \alpha}{v} \quad \text{and} \quad \frac{dt}{dx} = p. \qquad \ldots\ldots(7)$$

It follows, therefore, that the slope of the time-distance curve at any point gives the parameter of the ray passing through the point. In many cases the observations in the field have been found compatible with certain simple and continuous variations of velocity with depth, and for these cases

$$x = \int_0^z pv\, dz/(1 - p^2v^2)^{\frac{1}{2}},$$

$$t = \int_0^z dz/v(1 - p^2v^2)^{\frac{1}{2}} \qquad \ldots\ldots(8)$$

and

$$p = (\sin \alpha)/v = (\sin \alpha_0)/v_0 = dt/dx, \qquad \ldots\ldots(9)$$

where t is the time taken for the wave to travel from the origin (the shot point) to the point (x, z), z being positive in a downward direction, v_0 and α_0 being the surface value of the velocity and the angle between the ray and the vertical. If a horizontal reflecting boundary occurs at a depth h, the introduction of a factor 2 in the above expressions gives the reflection time T to a surface point distance X from the shot point, the upper limit of integration being h. Only in certain simple cases are the above integrals readily evaluated in a form suitable for computations, and the two most important velocity laws that have been found to fit field conditions closely are (i) the linear depth law $v = v_0 + az$, and (ii) the exponential law $v = v_0 e^{az}$. A continous increase of velocity with depth leads to a set of curved rays which are concave upwards, a fact which follows immediately.

since $\sin\alpha$ must increase with depth. All rays, except the special case $p=0$, penetrate to a maximum depth given implicitly by the equation

$$p=1/v,$$

and then return to the surface by a symmetrical path. The two cases cited above have been investigated mathematically by Ewing and Léet (1932b), by Slotnick (1936a, 36b), and by others. For the linear increase of velocity with depth, the rays are circles passing through the centre of disturbance and with their centres on the plane $z=-v_0/a$, i.e. the plane of zero velocity; the wave-front is an expanding sphere whose centre moves continuously downwards along the vertical through the shot point. For a ray which suffers no reflection or refraction at a velocity discontinuity

$$t=\frac{2}{a}\sinh^{-1}\frac{ax}{v_0}, \qquad \ldots\ldots(10)$$

where t is the time taken by the disturbance to reach a point on the surface a distance x from the shot, this being the time distance for the "direct" wave. For a wave reflected from a horizontal boundary at a depth h

$$\frac{at}{2}=\cosh^{-1}\frac{v_0}{2v_h}\left\{1+\left(\frac{v_h}{v_0}\right)^2+\left(\frac{ax}{2v_h}\right)^2\right\}, \qquad \ldots\ldots(11)$$

where v_h is the velocity at the greatest depth of penetration.

By a suitable selection of scales, a set of curves can be constructed which are applicable to all values of v_0 and a. Thus, using the substitutions

$$\tau=at/2, \quad \sigma=ax/v_0 \quad \text{and} \quad \lambda=v_0/v_h, \qquad \ldots\ldots(12)$$

a set of dimensionless quantities, the time equations for the two rays become

$$\tau=\sinh^{-1}\sigma. \qquad \ldots\ldots(13a)$$

and

$$\tau=\cosh^{-1}\frac{\sigma^2+1+\lambda^2}{2\lambda} \qquad \ldots\ldots(13b)$$

respectively. For reflection prospecting a series of curves showing the variation of τ with σ are drawn for different values of λ. From the observed values of t and x for the reflected pulse the values of τ and σ can be calculated, since v_0 and a are known, and from the chart the appropriate value of λ can be read for these co-ordinates. The depth follows immediately from the equation

$$h=v_0/a(\lambda-1).$$

Similarly, for the exponential velocity increase with depth, if

$$\sigma=ax/2, \quad \tau=av_0t/2 \quad \text{and} \quad \lambda=v_h/v_0=e^{ah}, \qquad \ldots\ldots(14)$$

then, for the direct wave,

$$\tau=\sin\sigma, \qquad \ldots\ldots(15a)$$

and for the reflected wave

$$\tau^2=1+1/\lambda^2-\frac{2\cos\sigma}{\lambda}. \qquad \ldots\ldots(16a)$$

Universal curves can again be drawn and the depth determination from the observed quantities is carried out in a similar manner.

In addition to these depth variations of velocity, others have been suggested and applied. Thus, for example, Mott-Smith (1939) points out that few velocity-depth laws lead to readily integrable expressions and, in addition, the two above, which are usually employed, imply that the velocity increases continuously, a phenomenon which cannot be reproduced in nature. Further, in the determination of v_0 and a, experimental time-distance curves, or time-depth curves in the case of well-shooting, have to be fitted to complex equations which do not permit of the easy application of least-square methods. The adoption of a formula giving a linear variation of velocity with time overcomes many of these objections, for in well-shooting the average velocity gives a linear relation with depth, and this can be fitted to the observations readily. Houston (1939) has shown that this assumption is equivalent to the parabolic depth relation

$$v = v_0(1 + kz)^{\frac{1}{2}}, \qquad \ldots\ldots(17)$$

and has developed a method whereby the time-distance curves for the direct wave and the reflection time-distance curves can be determined graphically. Another velocity relation was proposed by Muskat (1937), who expressed the depth as a polynomial of the velocity, i.e.,

$$y = (v - v_0)\Sigma a_n v^n, \qquad \ldots\ldots(18)$$

leading to forms which can be readily integrated. In addition, the formula

$$y = a \log (v_1 + v)(v_1 - v_0)/(v_1 - v)(v_1 + v_0) \qquad \ldots\ldots(19)$$

was proposed at the same time, since it leads to a velocity which has the velocity v_0 at the surface and increases asymptotically to the value v_1, a physical condition more probable than the continuous linear or exponential increase. Similar properties are also possessed by the form

$$y = a\left(\frac{v - v_0}{v_1 - v}\right) \qquad \ldots\ldots(20)$$

(v) *Determination of the velocity-depth relation*

The actual problem which confronts the prospector before any of these formulae can be applied is the experimental determination of the velocity variation, and in general three methods are available.

(i) *From time-distance curves.* The simplest is to determine the time-distance curve for the direct wave by using a series of geophones placed in a straight line at various distances from the shot point, also on the line. By comparing the observed curve with those calculated for the various assumed depth-velocity relations, it is often possible to obtain a numerical formula which will fit the observations closely. Certain time-distance curves, however, cannot be made to fit the derived formulae satisfactorily, and in this case a modification of the Herglotz-Bateman equation (Bateman, 1910), due to Ewing and Leet (1932b), can be employed. From simple considerations the slope at any point of the time-distance curve gives the velocity at the greatest depth reached by the ray, for, if v_0 is the surface velocity and the ray emerges at an angle α_0,

$$dx/dt = v_0/\sin\alpha_0 = v_{max}/\sin 90° = 1/p.$$

Consequently, if the depth of penetration can be determined, the velocity-depth variation is known. The modified formula gives

$$z = \frac{1}{\pi}\int_0^X \cosh^{-1}\frac{T'(x)}{T'(X)}\,dx, \qquad \ldots\ldots(21)$$

where z is the penetration of the ray emerging at a distance X and $T'(x)$ is the slope of the time-distance curve. Accordingly, by performing a series of graphical integrations, the various values of z can be ascertained.

(ii) *From well-shooting data.* Where possible, well-shooting gives more exact and reliable data on which to base the interpretation. In well-shooting, the geophone is lowered at the end of a cable to various depths in the well, and the time of transit from a surface explosion to the geophone is measured. Except for geophone depths which are shallow compared with the distance of the explosion from the well, in which case the ray path is appreciably curved, the results immediately give a close approximation to the average velocity to various depths. In addition, by measuring the interval velocities, i.e. the average velocities calculated from the time difference for two consecutive geophone positions, or the slope of the time-depth graph, the shape of the depth-velocity curve can be estimated. Since, for practical reasons, the explosion must take place some distance from the well, the time-depth curve exhibits a minimum time, corresponding to the ray which reaches the geophone horizontally. Here the above methods may lead to considerable error, but Dix (1939) has examined this aspect of the interpretation of well-shot data, and has developed a method which allows the depth variation to be calculated closely for any type of curve. Obviously, in those regions in which the velocity follows a mathematical law the interpretation becomes a determination of constants by comparison of the observed results with a theoretical curve.

(iii) *The velocity-profile method.* Both these methods have severe limitations. In the first, for deep penetration, the distance between shot point and receiver is very great and the uniformity of conditions becomes doubtful. More important, however, is the fact that, unless the velocity always increases with depth, there may be multiple arrivals in certain regions, those waves which have penetrated through the region of decreasing velocity sometimes arriving later than those which have not penetrated so deeply. With most refraction profiles, the interpretation is usually based on first arrivals, and hence there is a grave danger that the region of decreasing velocity may be undetected. The second method depends on the existence of a well passing through a representative rock sequence, near the site of the survey, and of a depth suitable for the examination of the velocity to a depth greater than the reflecting boundaries. Frequently this last condition is not satisfied, reflections being obtained from very deep-seated horizons, and for these regions efforts have been made to measure the average velocity using the data supplied by the reflections. The possibility of measuring the velocity in this manner has been appreciated for some time, and Rutherford (1934) attempted to do so, assuming the velocity to be uniform above the reflecting horizon. By using equation (1) for the reflected time for a horizontal boundary and obtaining a large number of corresponding values of t and x, he treated the system as a linear equation between t^2 and x^2

and solved them by least-square methods for v and p. He states: "The writer has solved a great many of these systems, ... but has never found a profile where the times could be determined to the degree of accuracy necessary for reliable results". Pugh (1934), however, pointed out that Rutherford used separate mechanical seismographs, each with its own independent time scale, and whose relative time accuracy cannot be compared with that of the central recording system employing electrical geophones and the same time base for all instruments. In addition, a modification of the problem before the application of least-square methods was suggested by Pugh. A first approximation is made for v_0 and p_0 from the data obtained from two geophones (usually the nearest and farthest instruments from the shot point), and using these values, a linear equation between the differences $(v-v_0)$ and $(p-p_0)$ is formulated using Taylor's theorem, where v and p are the true velocity and depth. Least-square methods are then applied to determine these corrections. Should the corrections be large, so that the use of a finite number of terms in the expansion is not justified, a second calculation is carried out using the modified values as a second approximation to the true values. The application of least-square methods to the seismic reflection problem has been developed more fully by Roman (1932), and the treatment by Pugh is a special case of the more general problem examined by him.

In spite of the criticism by Pugh, Green (1938) has applied the method described by Rutherford with considerable success, his experiments resulting in the "velocity-profile method". By selecting the site of the experiment at a point where there are several reflecting horizons, which must persist horizontally to give reflection records at all geophone positions, where the weathering is uniform and where the level differences at the surface are small, velocities have been measured with an accuracy of 3 per cent using geophone distances up to 4000 feet from the shot point. All shots were fired at the same depth, and since the formula applies only to horizontal reflectors, the region must be chosen so that this condition is fulfilled, or the line joining the shot to the geophones must be parallel to the strike of the beds. Green compared his results with those obtained using well-shooting data, the wells being 15 to 18 miles from the site of the velocity-profile experiments, and the two sets agreed to within 2·5 per cent. In these experiments the velocity was calculated from the slope of the line obtained by plotting X^2 against T^2. Such experiments are of importance in reconnaissance surveys in new areas where no wells are available. Thus, in Venezuela, a survey was made in the District of Infante, State of Guarico (Ransome and Romberg, 1941), and the velocity determinations were made both from refraction- and reflection-velocity profiles. Two years later, in the well Guayabo No. 1, direct observations on the velocity were made by well-shooting. The results of the refraction profile gave the smoothest velocity-depth curve, which agreed to within 100 feet per second (in velocities ranging from 7000 to 10,000 feet per second) with the values derived from the reflection data, except at a depth of 5000 feet. There was, however, a considerable discrepancy between the data obtained from the refraction and reflection experiments and the well data; the direct observations gave values some 500 feet per second greater over the major depth range, but close agreement was obtained at a depth of 4000 feet.

The velocity-profile method requires a special lay-out involving a large geophone spread, and for these the straight path and mean velocity assumptions may not be sufficiently to close the actual conditions. Romberg (1941) has proposed, and used, the results of routine field surveys for the determination of the velocity, using the relation

$$V^2 = \frac{\Delta X}{\Delta T} \cdot \frac{2X + \Delta X}{2T + \Delta T}, \qquad \ldots\ldots(22)$$

where T and X refer to the nearest geophone and $(T+\Delta T, X+\Delta X)$ to the far detector. Assuming that the same velocity variation applies at all points within the area, the total times are divided into groups consisting of consecutive intervals of 0·1 second, and the average T and ΔT are determined for each group and used to determine both the velocity and the depth. In the velocity-profile method the accuracy is increased by using a large value for ΔX and measuring a correspondingly large value for ΔT. In the last method, the "Delta-T" method, the accuracy is obtained by using a large number of observations. The main error in the method depends on the error in ΔT, and these will be caused by (i) the dip of the reflecting beds, (ii) irregularities in the weathered layer,* (iii) frequency changes with distance of the reflection record (the reflected pulses are usually observed at some phase later than the actual onset), and (iv) errors made in reading the record. If all stations are shot in both directions, (i), (ii) and (iv) will give random errors, which can be assessed, assuming a Gaussian distribution, but (iii) will be systematic, since the frequency of the reflected pulse will change continuously with distance. Its presence can be detected by measuring the velocities with two geophone spreads. With this method, the probable errors increase with depth, since ΔT decreases. To some extent this can be offset by enlarging the time interval of the group, or by using overlapping groups, the latter naturally leading to a smoother mean velocity-depth curve. The method has been applied by Steele (1941) to areas in the Texas and Louisiana Gulf Coast, and in all cases examined it was found that the Delta-T velocities were less, by from 3 to 9 per cent, than the velocities observed directly using well-shot data. Steele demonstrated that the effect of the curvature of the ray is negligible for the particular geometry of the surface arrangement employed and for the conditions obtaining in the areas examined. In addition, he showed that the error, arising from the procedure of timing the reflected arrivals about a cycle later than their actual arrival, was only about 1 per cent or less. The discrepancy, however, can be accounted for by assuming multiple reflections. From the known well-shot data the $T-\Delta T$ values can be calculated for a single reflection, and also for a wave which has been reflected twice at the boundary and once at the free surface. A graph of the observed values of T and ΔT lies between the graphs for these two cases, suggesting that the mean values are made up partially of single reflections and partially of multiple reflections. If this is the case, many reflecting horizons, calculated on the basis of a single reflection, have no physical significance.

* See § 2 (vii).

(vi) *Practical methods of calculation*

When the velocity variations have been determined by the methods outlined above, the depth and dip calculations can be carried out. If the velocity approximates closely to any of the simple mathematical relations, the appropriate formula is employed for the depth estimation. The simplest way in this case is to obtain the time T_0 for the normal ray, i.e. the ray which is reflected at normal incidence back to the shot point, this time being obtained by interpolation from times at geophones on each side of the shot point. This value is used in the depth determination. The dip is obtained by measuring the apparent surface velocity v_α of the reflected pulse, for

$$v_\alpha = v_0/\sin\alpha_0 = \Delta X/\Delta T = 1/p, \qquad \ldots\ldots(23)$$

where, as before, p is the parameter defining the ray. For the normal ray the angle between the ray and the vertical at the point of reflection is obviously the angle of dip θ, and

$$v_h/\sin\theta = \Delta X/\Delta T, \qquad \ldots\ldots(24)$$

an equation which allows θ to be calculated, since v_h is known at the measured depth.

Graphical methods can be used also to obtain the co-ordinates of the reflecting point. Thus, in the case of a linear increase of velocity with depth, for which the rays are arcs of circles and the wave-fronts spheres, a chart (Petersen, 1941), consisting of the wave-fronts $t_0 = \text{const.}$ and the rays $p = \text{const.}$, is superimposed on a rectangular co-ordinate system representing depths and horizontal displacements from the shot point. From the measured values of t_0 and ΔT (giving the parameter p of the ray), the point of reflection for the normal ray can be picked out on the curved co-ordinate system, and its co-ordinates read directly on the Cartesian mesh. Again the dip is obtained as before, or the tangent to the wave-front at the reflecting point employed. As Petersen points out, however, the continuous use of the two superimposed co-ordinate systems is extremely trying to the computer. To overcome this objection, he noted that the wave-front and ray equations in a vertical plane passing through the shot point are identical with those of the equipotential surfaces and lines of force due to two charged circular and parallel cylindrical conductors. By applying a suitable conformal transformation, the rays become straight lines radiating from a point O (figure 4), and the wave-fronts become circles centred on the same point. The earth's surface, $z = 0$, is transformed into a circle, the diameter passing through O, giving the direction of the vertical; equal depth lines transform into circles touching the circle $z = 0$ internally at a point P diametrically opposite to O. In addition the lines, $x = \text{constant}$, become circles tangential to OP at P. The angle between the transformed ray path and the vertical is equal to the angle of emergence α_0. In use an arm, carrying a graduated time scale t_0, is rotated about O through the estimated value of α_0, which is read on a suitably placed scale, and the depth and horizontal displacement of the reflecting point obtained from the co-ordinates of the point given by the two sets of circles

x = constant, z = constant. Based on this transformation, Wolf (1942) has developed a mechanical device for the computation of seismic paths when the velocity is a linear function of the depth.

In applying the formula for the dip, it is assumed that over the geophone spread the arriving wave is plane, while in actual fact the wave-front is part of a sphere. The measurement ΔT accordingly includes a certain time difference arising from the curvature. A simple correction has been developed by Widness and Haskell (1940) which, for a symmetrical geophone spread, takes the form

$$\frac{\Delta T}{X} = \frac{\sin \alpha_0}{v_0}\left\{1 - \frac{x^2}{8\rho^2}\cos^2 \alpha_0\right\}, \qquad \ldots\ldots(25)$$

where ρ is the radius of the wave-front. Since the second term in brackets

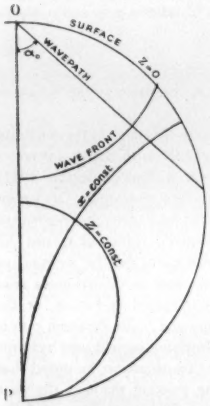


Figure 4. Transformed ray and wave-front chart. (*After Peterson*, 1941.)

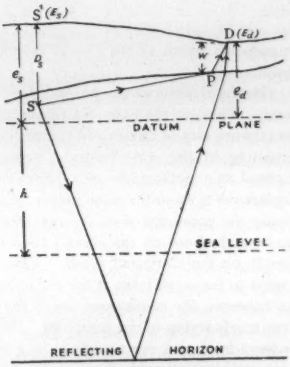


Figure 5. The calculation of surface corrections.

is a small correction, ρ may be taken as twice the computed depth, and, unless very steep dips are in question, $\cos^2 \alpha_0$ becomes unity. The same paper also investigates the effect of the curvature of the reflecting surface on the curvature of the emergent wave-front, and the conclusion is reached that, except in very rare cases, the effect can be ignored.

In many cases it will obviously be impossible to represent the velocity variations as a simple function of the depth in a form suitable for the direct integration of equations (8) for the time and depth. For this general case Waters (1941) has demonstrated how the interpretation may be carried out graphically. A

simple transformation of these equations leads to

$$T = 2\int_0^H dz/v^2(1/v^2 - 1/v_a^2)^{\frac{1}{2}} = 2\int_0^H F_1\,dz$$

and

$$R = \int_0^H dz/v_a(1/v^2 - 1/v_a^2)^{\frac{1}{2}} = \int_0^H F_2\,dz, \qquad \text{.....(26)}$$

where T is the travel time of the reflected pulse, R the horizontal displacement of the reflecting point from the shot point and v_a the apparent surface velocity, defined in equation (23). For a series of suitable values of v_a, corresponding to various values of ΔT, these integrations can be performed graphically, and a series of curves compiled showing the relation between T and H and R and H for various values of $\Delta T/X$. From the observed values of T and ΔT the depth is obtained directly from the first set of curves, and using this value, R is obtained from the second set, the correct curve to employ being fixed by the measured value of $\Delta T/X$. As before, the dip is obtained from equation (24).

The application of these methods requires that the direction of shooting coincide with the direction of dip; in practice, however, this latter direction is unknown. The general problem of measuring the direction of dip, as well as the dip, can be accomplished by using two intersecting lines of geophones. The three-dimensional problem has been examined by Rock (1938) and a nomogram for assisting the interpretation has been given by Lawlor (1938).

In certain areas where there is great surface erosion, the direct use of these methods of depth calculation often leads to results which show evidence of the influence of surface topography, valleys being revealed by subsurface ridges and surface elevations by subsurface valleys. Examples of these have been found in western New York, Pennsylvania and in Wyoming, where surface-level difference of 1000 feet may be encountered (Beers, 1939). Such errors arise from the assumption that v_0, the surface velocity, is constant over the area, for obviously the surface velocity at the lower points corresponds to a rock layer well below the surface in the elevated regions, and the effective value of v_0 must be selected accordingly.

(vii) *The weathered layer and the surface correction*

The observed values obtained from the seismograms require certain corrections before they are used in the above methods of interpretation, the corrections arising from (i) the depth at which the explosion occurs, (ii) the differences in the elevations of the various geophones and (iii), most important of all, the thickness of the surface layer, usually referred to as the weathered layer, and its irregularities. The existence of this surface layer has been appreciated for a long time, and it is characterized by an abnormally low velocity of propagation of the P waves in the layer. Little systematic work has been carried out to investigate the cause of these abnormal velocities, but it is found that the layer is almost universal and its thickness varies from a few feet to as much as 200 feet; actual examination of the rocks reveals no visible difference between the layer and the adjacent underlying material (Lester, 1932). Nevertheless, velocities as low as 2000-2500 feet/second have been observed, while the rock immediately

below it may transmit the waves with a velocity of 5500 feet/second. It should be noted that the velocity change is sharp, and not a gradual transition. The name given to the layer is rather unfortunate, since it does not agree with the zone of geological weathering, but in a large number of cases it appears to be bounded by the water horizon. In some cases (Lester, 1932) the observed velocity was less than that of sound in air, and Lester suggested that the low velocities could be explained on the basis of air inclusions. so that the system was similar to the case of a water-air mixture examined by Wood (1930). Basing the calculation on a formula given by Wood, velocities much lower than that of sound in air are possible. The suggestion is further strengthened by the fact that in certain layers the velocity decreases towards the surface, where a greater air content would be anticipated.

The correction for this layer cannot be treated as part of the velocity variation with depth because of the frequently irregular thickness of the layer over the surface lay-out, i.e. in this region the velocity is not a function of depth only. The methods employed for this correction, together with those for the level difference between the shot point and the geophone, have been summarized by Nettleton (1940). The total correction, usually called the surface correction, reduces the observed time to the time which would have been taken by the wave if the surface were horizontal and the weathered zone absent. It can be made to different datum planes, for example to the elevation of the surface at the shot point, the elevation of the shot itself, or to some arbitrary elevation. No additional experiments are required, the first arrivals at the various geophones giving the data from which the thickness of the weathered zone can be computed, using those principles developed for the interpretation of refraction shooting. The shot is usually placed below the weathered zone, for it is found that here a more efficient energy transfer to the ground takes place than in the weathered zone itself. The first arrival is a wave which has followed a refraction path SPD (figure 5), and its arrival time after the explosion is given approximately by

$$t = \frac{x}{v_0} + \frac{W \cos i}{v_1}, \qquad \ldots\ldots(27)$$

where x is the horizontal distance from shot to geophone, v_1 and v_0 the velocities in the weathered zone and the underlying rocks respectively, i the critical angle, and W the thickness of the weathered layer at the geophone. Accordingly, the arrival times are plotted against distance, and the slope of the best straight line through them gives the value of v_0. Alternatively, since v_0 is the effective surface velocity, this value is usually known. Through the origin a line of the same slope is constructed ($t = x/v_0$), and the time difference t between this line and the observed arrival times can be associated with the last term in equation (27), and from this W may be determined for each geophone.

To indicate the method of obtaining the correction the following gives the surface correction to an arbitrary datum plane at an elevation h above sea-level. If e_s is the surface elevation of the shot point above the datum plane, e_d that of

the detector, the correction consists of

(i) correction for time from shot to datum plane $= -\dfrac{e_s - D_s}{v_0}$;

(ii) correction for time from datum plane to base of weathered layer $= -\dfrac{e_d - W_0}{v_0}$;

(iii) correction for time in weathered layer $= -\dfrac{W}{v_1}$;

where D_s is the depth of the shot from the surface. The total surface correction is

$$c_s = -(e_s - D_s + e_d)/v_0 - W(1/v_1 - 1/v_0)$$
$$= -(e_s + e_d - D_s)/W - t'\{(v_0 - v_1)/(v_0 + v_1)\}^{\frac{1}{2}}. \quad \ldots\ldots(28)$$

The actual observed values are E_s and E_d, the elevation of the surface of the shot point and the detector above sea-level, and finally

$$c_s = -(E_s + E_d - 2h - D_s)/v_0 - t'\{(v_0 - v_1)/(v_0 + v_1)\}^{\frac{1}{2}}. \quad \ldots\ldots(29)$$

It will be noted that the angle between the ray and the vertical does not appear in these formulae, and it is implied that the angle is always small as, indeed, it will be for reflections from deep interfaces. This may not be true for shallow reflections, and errors may then arise. After the correction has been applied, the depth calculated naturally gives the distance below the datum plane. The importance of the weathering correction can easily be appreciated from the fact that "a difference of 3·5 feet in the thickness of the low-velocity zone between the ends of the geophone spread may easily produce an error of about 1° in the computed dip if not properly corrected" (Rosaire and Adler, 1934).

§ 3. GRAVITATIONAL AND MAGNETIC METHODS

In the gravitational method the main interest has been focused on the design and field use of gravity meters. As is well known, the early application of this method made use of the Eötvös torsion balance, an instrument for the measurement of the horizontal gradient of gravity and the differential curvature of the level surfaces. This instrument suffers from two main defects: it is slow in operation and, more important, it is influenced greatly by surface irregularities in its immediate vicinity. The modern static gravity meters are far quicker to use and they are not so sensitive to the surface topography, and accordingly may be used in areas of greater surface relief than the torsion balance. On the other hand, the gravity meter is responsive to large deep-seated masses and is influenced by those density variations associated with the isostatic compensation of mountains, etc., and giving rise to a regional gravity anomaly. Thus the new type of instrument cannot be regarded as displacing the torsion balance, for the two are, to a large extent, supplementary, the Eötvös torsion balance being suitable for detailed work and small shallow density anomalies, and the gravity meter for large deep-seated structures and for reconnaissance work.

(i) *Modern gravity meters*

To a limited extent pendulums have been employed for prospecting purposes (Gay, 1940), but the static gravity meter is far more sensitive, being capable of detecting changes in gravity as small as 0·1 milligal (Bruckshaw, 1941), a fractional change in gravity of 1 in 10^7. To attain this sensitivity a number of precautions must be taken in the design and construction of the instruments. Practically all gravity meters consist of an elastic member sustaining the gravitational force acting on a fixed mass. Two main types exist, those in which Hooke's law is applied directly, the fractional change in the displacement of the mass being equal to the fractional change in gravity, and those consisting of a spring-controlled mass system operating near a position of instability, a method adopted to obtain increased sensitivity. No matter what system is utilized, the important feature is the spring, whose elastic properties will change with temperature and, in addition, it will exhibit, to some degree, the defects arising from elastic hysteresis. Accordingly, a limit to the accuracy of any gravity meter is imposed by the imperfections of the material of the elastic member, and few substances are suitable; of these, quartz, certain steels (elinvar) and gases are notable examples. In most cases the material selected has a low temperature-coefficient of elasticity; but, even so, temperature compensation is utilized in addition to controlling the temperature of the instrument. This latter feature is more important than the compensation, for unless the temperature over the whole instrument is constant, the differential expansion between the parts vitiates the readings (Hartley, 1932). Most gravity meters show a gradual drift due to ageing and slow yield of the spring, but in a well-constructed instrument this is a regular change with time, and can be corrected by a repetition of stations at different times (Bryan, 1937; Nettleton, 1941).

Many gravity meters have been developed by different investigators, as, for example, the Hartley instrument (Hartley, 1932), employing a simple stretched spring whose length is maintained constant by a light micrometer-operated "weighing" spring, the Haalck instrument (Haalck, 1932), and those due to Nørgaard (1933, 35, 36) and Benfield (1938), all utilizing an enclosed mass of gas as the elastic member. Of those operating near a position of equilibrium, the Ising and Urelius gravity meter (1930) and the Holweck-Lejay instrument (1934) may be cited. To indicate the trend of modern design, and the methods of obtaining the necessary magnification, the Boliden instrument and the Gulf gravity meter, employing electrical and optical magnification respectively, are here described.

The former (Sundberg, 1938) consists of a mass (figure 6) supported by two bent springs from the instrument frame, and each end of the mass forms an electrical condenser with a metal plate insulated from, but rigidly attached to, the frame. If a gravity change Δg produces a displacement Δx of the mass, then the change ΔC in the capacity C of the condenser is given by $\Delta C/C = -\Delta x/t$, where t is the separation of the plates. Thus, by making t sufficiently small, considerable magnification is possible. In the original paper it is claimed

that a mass displacement of $3{\cdot}5 \cdot 10^{-9}$ cm. can be detected, while a change of 1 milligal causes a displacement of $5{\cdot}5 \cdot 10^{-7}$ cm. Thus, theoretically, a gravity difference of 0·02 milligal can be detected, but it is doubtful if this accuracy can be attained. The condenser forms part of an oscillating circuit whose frequency changes with capacity, and the measurement is essentially one of comparing the frequency with that of an independent oscillator. It can also be used as a null instrument by utilizing the electrostatic attraction between the plates to restore the frequency to a standard value. The same method can also be employed to calibrate the instrument. An additional method of calibration is by tilting the instrument, and so employing only a component of gravity, since the instrument reveals only those displacements perpendicular to the plates. By suitable selection of materials (not specified) the temperature effect is made small, 1° c. corresponding to 2 milligals, while the temperature of the instrument is thermostatically controlled to 0·02° c. It is interesting to note

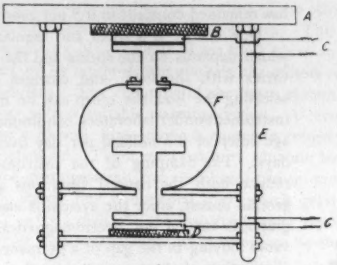


Figure 6. The Boliden gravity meter.
(*Reproduced, by kind permission, from Bull. Inst. Min. Met., Lond.*, no. 402, 1938.)

the purpose for which the instrument was developed. In electrical prospecting for good conductors, such as certain base metal sulphides, a number of uneconomic deposits, clays, graphitic schists, etc., give similar indications and, in general, there is no electrical method of differentiating between the two types. It was anticipated that a sufficiently sensitive gravity meter would allow a separation of the important and unimportant indications on account of the greater density of the sulphides.

The Gulf gravity meter (Wyckoff, 1941) was developed by the Gulf Research and Development Co. of Pittsburgh, Pa., and by 1936 had replaced completely the pendulum equipment previously used by that company. The instrument consists of a simple spring supporting a mass and an optical magnification of about 1000 between the mass and the index point. The spring takes the form of a ribbon helix with the greater dimension of the cross-section parallel to the

axis of the helix (figure 7). With such a spring a rotation of one end relative to the other occurs on loading, and the initial load of the gravity meter is such that about 8 complete revolutions take place, the spring actually uncoiling. The change of gravity is determined by observing the additional rotation of the lower end, this rotation being observed by means of a mirror attached to the mass and a multiple reflection system. Using an optical lever of 22 inches and observing the fourth reflection, 0·1 milligal corresponds to a displacement of 0·001 inch in the fiducial mark. The index is observed by a micrometer microscope, and actually a tenth of this displacement can be read. This simple system has many advantages, for it can easily be adjusted to any gravity range by a rotation of the upper end of the spring, which is attached to a torsion head. Further, the scale value of the instrument is unaffected by this adjustment, and the range of the microscope allows a change of 30 milligals to be observed on one adjustment. Tests show that over several years the calibration of the same instrument has remained constant to 0·5 per cent or better.

Figure 7. Spring system of the Gulf gravity meter. (*Reproduced, by kind permission, from* '*Geophysics*', 1941.)

One feature is a small and regular creep, a feature which depends on the spring and the load. The creep varies with the load, and changes sign. Thus, by selecting the load the creep can be made small. One instrument under laboratory conditions showed an average creep of 0·3 milligal per day over a period of 30 days. The damping of the instrument, necessary to reduce both the natural vibrations and those due to ground unrest, since the system is essentially a seismograph, is carried out electromagnetically by a series of vanes moving in the gap of a permanent magnet. The elimination of short-period ground motion is naturally assisted by the long periods of the system, the longitudinal period being about 1 second and the rotational period 7 seconds. Since the moving system has only one point of contact with the frame, it is relatively insensitive to tilt. There is, however, the bending of the spring near the support, and this is converted into a rotation of the helix end. By tapering the helix near the support, this effect is substantially reduced. The amount of tilt required to produce an error of 0·1 milligal varies considerably from about 10 seconds of arc in the worst case to over 60 seconds in the best instruments.

The reflecting flats in the multiple reflection system form parts of plano-convex lenses to avoid loss of light, and the partially reflecting surfaces have a transmission coefficient of 12 per cent to give maximum intensity to the fourth reflection. The clamping mechanism is so designed that the system is practically undisturbed from its equilibrium position by its operation. Although the

temperature coefficient of the spring is small, and temperature compensation is incorporated, temperature control is employed, and under field conditions this is effective to within 0·02° c. In operation it has been found possible to measure 100 stations, separated by one-mile intervals, in an eight-hour period, the work requiring two people. The normal speed is, however, about half this. As to accuracy, analysis of repeated stations, measured during routine field surveys, shows that the probable error of a single observation is less than 0·05 milligal.

An interesting development of the Gulf instrument is its application to underwater surveys (Pepper, 1941). The two main problems to be solved for this purpose involved the accurate levelling and the recording of observations. The former is carried out by remote control. The instrument, suitably housed, carries two mutually perpendicular bubble levels whose positions are observed by means of photo-electric cells. Light from a lamp is reflected, by means of two mirrors, past the ends of the bubble on to two photo-cells joined in parallel, with the positive of one cell joined to the negative of the other, the combination being joined to a micro-ammeter with a central zero. This system was found to give a level sensitivity of 3 seconds of arc for 1 micro-ampere. The actual levelling is carried out by motor-driven gimbals, the motors running at 10 r.p.m. for the fine adjustment. The clamping of the beam system is also carried out by electric motors with a suitable warning device. A record is taken of the beam position on a 35-mm. film, giving a complete record from the time the instrument is unclamped. Fixed marks from a separate lamp are also photographed, to allow compensation for shrinkage during processing. In quiet water a perfectly steady line is obtained from which gravity changes can be readily estimated to 0·02 milligal. In disturbed water the beam system never attains a steady state, but, nevertheless, its position can be observed to 0·1 to 0·2 milligal.

Under laboratory conditions modern gravity meters can attain a high sensitivity, as indicated by figure 8, which shows a set of observations at a fixed point and reveals the variations due to the effects of the moon and sun, etc. The circles represent the observed values, taken on an astatized instrument, plotted against time, the instrument employed being again designed by the Gulf Research and Development Co. Both this instrument and the one just described were developed simultaneously, but, for field use, the former was abandoned in favour of the latter.

Gravity observations require correction for a large number of factors before the measurements can be utilized for geological interpretation. Thus, field stations to the north or south of the datum station require correction, since part of the observed gravity difference is due to the general, and well known, variation of gravity from the equator to the poles. In addition, corrections for height are necessary, part of this taking into account the variation of distance of the point from the earth's centre (free air correction), the other part being Bouguer's correction. In its simplest form the last ignores the immediate surface irregularities, and assumes that the surface is a plane through the observation station.

Thus the attracting mass below each station depends on the station elevation, the higher the station the greater the mass. Compensation is necessary for the excess mass lying between the datum plane and the plane through the station, the correction being of the form kh, where h is the level difference and k a factor proportional to the density. The measurement of the representative density of a rock layer some hundreds of feet thick is difficult, and Nettleton (1939) suggested an indirect method of obtaining the correct value. From the form of the correction (kh), Bouguer's effect will superimpose on the other gravitational effects a part which reflects the height variations. To apply Nettleton's method a series of measurements is made along a straight line crossing some simple topographical feature, such as a hill. The feature should have gentle slopes

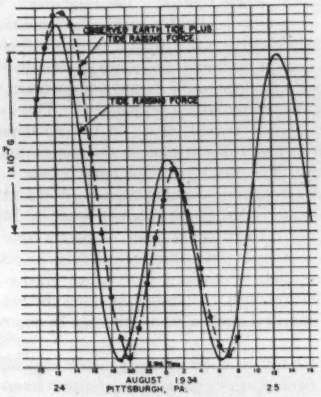


Figure 8. Observed time variation of gravity.
(*Reproduced, by kind permission, from* '*Geophysics*', 1941.)

so that they have little effect, and where possible there should be no gravitational anomaly due to buried geological structures. Under these conditions the measured values will exhibit variations corresponding to the surface elevations. The corrections are applied for a number of different densities and a set of curves obtained. Those curves for which the density employed is too small will still exhibit the surface variations but to a reduced degree, while a greater density will change maxima to minima and vice versa. The correct density will give a smooth curve showing no relation to the topography. Accordingly an inspection of the set of corrected curves allows the correct density to be selected. An improvement of this method has been put forward by Siegert (1942). This improvement is based on the calculation of k required to make the expression $\Sigma(g_i + kh_i)^2$ a minimum, where g_i is the excess gravity, and h_i the excess elevation, at any station calculated on the assumption that both vary linearly between

the two stations adjoining the observation point. With the earlier, and less sensitive, gravity meters the Bouguer correction was, in general, sufficient, but with the more accurate instruments the effects of the immediate local topography cannot be neglected. Hammer (1939) has published tables, modelled on those due to Hayford and Bowie (1912, 17), which allow this effect to be calculated with sufficient accuracy. It is worthy of note that, in addition to these corrections, Adler (1942) has prepared charts to correct gravity-meter surveys for the effect of the solar and lunar tidal forces.

(ii) *The interpretation of gravity observations*

The question of the interpretation of gravitational and magnetic anomalies, two closely related subjects (Gutenberg, 1926), has been examined theoretically by numerous investigators, including Barton (1929), Lancaster-Jones (1929), Nippoldt (1930), Shaw (1932), Pentz (1940), Fisher (1941), Levine (1941) and many others. Probably the most interesting development from the physical point of view is the experimental method developed by Fisher (1940) for computing magnetic and gravity effects due to bodies of irregular shape. The additional gravitational force Δg due to a horizontal lamina of density σ and thickness Δz is given by

$$\Delta g = G\sigma\Delta z \,.\, \omega,$$

where G is the constant of gravitation and ω the solid angle subtended by the lamina at the point of observation. Similarly, the magnetic potential Ω obtained when a wire, carrying a current i, coincides with the boundary is

$$\Omega = \omega i,$$

whence

$$\Delta g = \frac{G\Delta z\sigma}{i} \,.\, \Omega.$$

Hence, by measuring the variation of the magnetic potential, the gravity variation can be obtained by means of a numerical factor. Since a body of any shape can be divided into a number of horizontal laminae, the method can be applied in the general case by winding wire round a model of the structure, the wire following the contours. The variation of magnetic potential can be measured by a magnetic potentiometer, or by passing an alternating current through the wire contours and measuring the mutual inductance, by a standard method, between it and a long thin solenoid, one end of which coincides with the point of observation, the other end being well removed from the model. The method obviously allows the direct tracing of isogams or lines of equal gravity.

The method can also be extended to magnetic anomalies if it can be assumed that the body is uniformly magnetized. In this case the vertical magnetic anomaly ΔH_z due to the lamina is

$$\Delta H_z = k\left(H_x \frac{\partial\omega}{\partial x} + H_y \frac{\partial\omega}{\partial y} + H_z \frac{\partial\omega}{\partial z}\right) . \Delta z,$$

where k is the susceptibility and H_x, H_y, H_z are the components of the earth's

field. If H is the total earth's field and $\frac{\partial \omega}{\partial n}$ the resultant gradient,

$$\Delta H_z = kH\Delta z \frac{\partial \omega}{\partial n} . \cos\theta,$$

where θ is the angle between the field and the gradient direction. Thus it is only necessary to measure the field due to the current in the wire model in the fixed direction of the earth's field to obtain the variations of ΔH_z. For irregularly shaped bodies this method leads to a considerable saving in the time of calculation.

Of more general application is the discussion on the reliability of geophysical anomalies by Elkins (1940). All data contain random errors, and in the case of a small anomaly the question arises as to whether the anomaly is real, or due to an accidental arrangement of errors. To decide which of these is correct, it is assumed that the anomaly is a collection of random errors, and the probability that such a collection of random errors can occur is computed. If the probability is small, it is considered that it is unlikely that the anomaly is due to errors and is, in fact, a reflection of geological conditions. On the other hand, if the probability is large it is considered that the anomaly is valueless.

(iii) *The residual magnetism of rocks*

One novel application of magnetism to prospecting arises from the study of the residual magnetism of rocks, and the inferences that may be drawn from them. Measurements of residual magnetism, particularly of igneous rocks, have been carried out for a long time (Koenigsberger, 1938) and, recently, McNish (1937) has summarized the main ideas involved. In the case of igneous rocks it is assumed that the various mineral constituents, magnetite and ilmenite being the most important, acquire their residual magnetism in the earth's field existing at the time they cool below their Curie point. Similar fossil magnetism may have been imparted to the contiguous rocks suffering thermal metamorphism under conditions involving sufficiently large temperature changes. For sedimentary rocks, accumulating under quiescent conditions, as in lakes, the magnetized particles, settling slowly, align their axes parallel to the existing field. Thus an examination of the fossil magnetism may be expected to yield information concerning the direction of the earth's field, and possibly, in the case of igneous rocks, of the intensity in past ages. Before any reliable deductions can be made, the possibility of earth movement, chemical action, thermal metamorphism, and so forth, must be taken into account, in addition to the effects of mechanical shocks due to earthquakes, or magnetic shocks due, for example, to a lightning discharge. In addition, with the accumulation of sufficient data the geological dating of some types of rocks may be materially assisted. In this connection it is worthy of note that the magnetic properties of magnetites may have a certain diagnostic value in assessing their origin (Davis, 1936).

The magnetic mineral content, and the consequent fossil magnetism, are very small, and for these investigations an electromagnetic method (Johnson and McNish, 1938) has been developed which will allow small intensities of magnetization of $2 . 10^{-6}$ c.g.s. units/cm.^{3} to be measured with an accuracy of 2 per cent. The method consists essentially of rotating the rock specimen, in

the form of a cube, about a vertical axis, the centre of the specimen coinciding with the centre of a vertical coil. The e.m.f. induced in the coil is fed to a tuned, high-gain amplifier (Johnson, 1936), whose output passes through a micro-ammeter via a synchronous commutator. To prevent parasitic effects arising from fluctuating magnetic fields, or from forced vibrations of the coil, a second coil, of greater linear dimensions but of the same effective area, is mounted concentrically and parallel to the first. Electrostatic shielding also is necessary to reduce the effects of extraneous charges, particularly those due to static charges on the non-conducting specimen holder. Some idea of the sensitivity, and the importance of these precautions, is given by the fact that an angular vibration of 0·001 second of an uncompensated coil would give observable deflections. The observations were made by adjusting the commutator phase displacement until zero reading was obtained and then measuring the output with the phase displacement changed by $\pi/2$. For calibration purposes a small magnet of known moment was employed. It would appear, however, that the system is only suitable for uniform rocks of fine-grained texture. A large crystal of high susceptibility rotating about the axis would give a fluctuating field indistinguishable from that due to permanent magnetism.

The method has been applied (McNish and Johnson, 1938) to determining the direction of magnetization of varved Pleistocene clays from New England and of ocean-bottom sediments. The variation in field direction over a continuous sequence of years is possible with the varved clays, for they show a regular variation in texture due to seasonal deposition, and their changing thickness allows them to be placed chronologically on the Antevs scale (Antevs, 1928). The measurements ranged over the years 211 to 408, Antevs chronology, and the results exhibited a declination varying from 11 degrees to 42 degrees west, showing an average polarization nearly 30 degrees west of the present declination. The results from the marine sediments were not so satisfactory, since there was no time scale and the orientation of the specimen with respect to the present magnetic meridian was unknown. Nevertheless, in a thickness of 250 cm. the direction of polarization varied through a range of 70 degrees.

Observations of a similar nature have been applied to determine the orientation of cores obtained from deep wells in sedimentary deposits. The magnitude of the dip of beds penetrated by the drill can be obtained satisfactorily from their trace on the curved walls of the core but, due to the difficulty of determining the orientation of the core by mechanical means (Macready, 1930), the direction of dip is unknown. The direction of the polarization (Lynton, 1937, 38) is obtained by mounting the core, cut to suitable size and treated to remove surface contamination, with its axis horizontal and close to one pole of a shielded astatic magnetic pair suspended by a fine torsion wire. A continuous record of the deflection is taken as the specimen is turned slowly through one revolution, the deflection being partly due to residual polarization and partly to the mineral susceptibility. The latter, frequently irregular in coarse-grained rocks, is eliminated by reversing the axis and repeating the process. Contrary to the effects discovered by McNish and Johnson, it is assumed that the direction of polarization coincides with that of the earth's field. In this connection the

authors claim that, for tests on outcrop samples, the direction of polarization coincides, within a few degrees, with the existing earth's field. One other difficulty arises from the dip of the beds themselves, which implies earth movement since their deposition, and accordingly the residual axis of polarization, need not coincide with the earth's field at the present time or at the time of formation. The possibility of alteration due to weathering must not, however, be excluded, and at the same time it is possible that the physical and chemical processes undergone by deeper strata may lead to a direction of magnetization agreeing more closely with the present field, changes which the recent varved clays have not suffered.

Some rocks are unsuited to this process, particularly limestone, anhydrite, dolomite, etc., which contain little or no magnetic minerals. To overcome this and other uncertainties Vacquier (1939) has suggested the use of a cement, carrying polarized material, injected into a pilot hole in the formation to be examined, the core being extracted after the cement has set. Such a process would undoubtedly eliminate the main difficulties encountered, but considerably longer time would be required for this than for the Lynton method, which does not interfere with the drilling.

REFERENCES

ADLER, J. L., 1942. *Geophys.* **7**, 35.
ANTEVS, E., 1928. *Amer. Geog. Soc., Research Series*, no. 17.
BARTON, D. C., 1929. *Geophys. Prospecting*, 480 (New York: A.I.M.E.).
BATEMAN, H., 1910. *Phil. Mag.* **19**, 578.
BEERS, R. F., 1939. *Geophys.* **4**, 167.
BENFIELD, A. E., 1938. *Mon. Not. R. Astr. Soc., Geophys. Suppl.*, **4**, 351.
BORN, W. T., 1941. *Geophys.* **6**, 132.
BRUCKSHAW, J. McG., 1941. *Proc. Phys. Soc.* **53**, 449.
BRYAN, A. B., 1937. *Geophys.* **2**, 301.
DAVIS, C. W., 1936. *Economic Geol.* **30**, 655.
DIX, C. H., 1939. *Geophys.* **4**, 24.
ELKINS, T. A., 1940. *Geophys.* **5**, 321.
EWING, M. and LEET, L. D., 1932 a. *Geophys. Prospecting*, 245 (New York: A.I.M.E.).
EWING, M. and LEET, L. D., 1932 b. *Geophys. Prospecting*, 263 (New York: A.I.M.E.).
FISHER, J. W., 1940. *Geophys.* **5**, 22.
FISHER, J. W., 1941. *Geophys.* **6**, 1.
GAY, M. W., 1940. *Geophys.* **5**, 176.
GOLDSTONE, F., 1934. *Proc. World Petroleum Congress*, 155.
GREEN, C. H., 1938. *Geophys.* **3**, 295.
GUTENBERG, B., 1926. *Lehrbuch d. Geophys.* 557 (Berlin).
GUTENBERG, B., 1936. *Beitr. Geophys.* **6**, 125.
HAALCK, H., 1932. *Z. Geophys.* **8**, 17. 197.
HAMMER, S., 1939. *Geophys.* **4**, 184.
HARTLEY, K., 1932. *Physics*, **2**, 123.
HAYFORD, J. F. and BOWIE, W., 1912. *U.S.C. and G.S., Special Pub.* no. 10.
HAYFORD, J. F. and BOWIE, W., 1917. *U.S.C. and G.S., Special Pub.* no. 40.
HOLWECK, F. and LEJAY, P., 1934. *J. Observations*, **17**, 109.
HOUSTON, C. E., 1939. *Geophys.* **4**, 242.
HOWELL, L. G., KEAN, C. H. and THOMSON, R. R., 1940. *Geophys.* **5**, 1.
ISING, G. and URELIUS, N., 1930. *Bull. Géodésique*, Annexe 7, 556.
JOHNSON, A. E., 1936. *Physics*, **7**, 130.
JOHNSON, A. E. and MCNISH, A. G., 1938. *Terr. Mag. and Atmos. Elec.* **43**, 401.
JOHNSON, C. H., 1939. *Geophys.* **4**, 33.

KLIPSCH, P. W., 1936. *Geophys.* **1**, 365.
KNOTT, C. G., 1899. *Phil. Mag.* **48**, 64.
KOENIGSBERGER, J. G., 1938. *Terr. Mag. and Atmos. Elec.* **43**, 119, 299.
LANCASTER-JONES, E., 1929. *Geophys. Prospecting*, 505 (New York : A.I.M.E.).
LANCASTER-JONES, E., 1936. *Rep. Prog. Phys.* **2**, 97.
LAWLER, R., 1938. *Geophys.* **3**, 340.
LESTER, O. C., 1932. *Amer. Assoc. Petr. Geol., Bull.*, **16**, 1230.
LEVINE, S., 1941. *Geophys.* **6**, 180.
LYNTON, E. D., 1937. *Amer. Assoc. Petr. Geol., Bull.*, **21**, 580.
LYNTON, E. D., 1938. *Geophys.* **3**, 122.
MCDERMOTT, E., 1932. *Amer. Assoc. Petr. Geol., Bull.*, **16**, 34.
MCNISH, A. G., 1937. *Terr. Mag. and Atmos. Elec.* **42**, 283.
MCNISH, A. G. and JOHNSON, A. E., 1938. *Terr. Mag. and. Atmos. Elec.* **43**, 401.
MACREADY, G. A., 1930. *Amer. Assoc. Petr. Geol., Bull.*, **5**, 571.
MOTT-SMITH, M., 1937. *Geophys.* **2**, 265.
MOTT-SMITH, M., 1939. *Geophys.* **4**, 8.
MUSKAT, M., 1937. *Geophys.* **2**, 319.
MUSKAT, M. and MEERES, M. W., 1940 a. *Geophys.* **5**, 115.
MUSKAT, M. and MEERES, M. W., 1940 b. *Geophys.* **5**, 149.
NETTLETON, L. L., 1939. *Geophys.* **4**, 176.
NETTLETON, L. L., 1940. *Geophysical Prospecting for Oil*, 305 (London : McGraw-Hill Publishing Co.).
NETTLETON, L. L., 1941. *Geophys.* **6**, 270.
NIPPOLDT, A., 1940. *Verwertung magnetischer Messungen zur Mutung* (Berlin).
NØRGAARD, G., 1933. *Balt. Geodet. Komm.* 6 Tagg. 211.
NØRGAARD, G., 1935. *Balt. Geodet. Komm.* 7 Tagg. 279.
NØRGAARD, G., 1936. *Balt. Geodet. Komm.* 8 Tagg. 127.
PENTZ, H. H., 1937. *A.I.M.E. Tech. Pub.* 766 (New York).
PENTZ, H. H., 1940. *Geophys.* **5**, 295.
PEPPER, T. B., 1941. *Geophys.* **6**, 34.
PETERSEN, R. A., 1941. *Geophys.* **6**, 74.
PIRSON, S. J., 1937. *A.I.M.E. Tech. Pub.* 833 (New York).
PUGH, W. E., 1934. *Geophys. Prospecting*, 455 (New York : A.I.M.E.).
RANSOME, W. R. and ROMBERG, F., 1941. *Geophys.* **6**, 158.
RIEBAR, F., 1936. *Geophys.* **1**, 97.
RIEBAR, F., 1937. *Geophys.* **2**, 132.
ROCK, S. M., 1938. *Geophys.* **3**, 340.
ROMAN, I., 1932. *Geophys. Prospecting*, 460 (New York : A.I.M.E.).
ROMBERG, F., 1941. *Geophys.* **6**, 356.
ROSAIRE, E. E. and ADLER, J. L., 1934. *Amer. Assoc. Petr. Geol., Bull.*, **18**, 119.
RUTHERFORD, H. M., 1934. *Geophys. Prospecting*, 391 (New York : A.I.M.E.).
SALVATORI, H., 1937. *Geophys.* **2**, 342!
SHAW, H., 1932. *Trans. A.I.M.E.* **97**, 271.
SIEGERT, A. J. F., 1942. *Geophys.* **7**, 29.
SLOTNICK, M. M., 1936 a. *Geophys.* **1**, 9.
SLOTNICK, M. M., 1936 b. *Geophys.* **1**, 299.
SPECHT, Z., 1940. *A.I.M.E. Tech. Pub.* 1177 (New York).
STEELE, W. E., 1941. *Geophys.* **6**, 370.
SUNDBERG, K., 1938. *Bull. Instr. Min. Met., Lond.*, no. 402.
VACQUIER, V., 1939. *Geophys.* **4**, 292.
WATERS, K. H. 1941. *Geophys.* **6**, 34.
WEATHERBY, B. B., 1940. *Geophys.* **5**, 215.
WIDNESS, M. B. and HASKELL, N. A., 1940. *Geophys.* **5**, 116.
WOLF, A., 1942. *Geophys.* **7**, 61.
WOOD, A. B., 1930. *Textbook of Sound*, 327 (London : G. Bell and Sons, Ltd.).
WYCKOFF, R. D., 1941. *Geophys.* **6**, 13.

SOUND

By E. G. RICHARDSON

§1. PROPAGATION IN FREE MEDIA

Sieg (1940) and Eagleson (1941) have both done theoretical and experimental work on the propagation of sound in the open air. Wind and temperature gradients are, of course, the principal disturbing factors. With winds over 10 miles per hour it is the former factor which predominates, temperature inversions and gradients being only important in calm weather. Acoustically good weather is characterized by a calm, and accompanied by overcast sky or mist; bad weather by high temperature-lapse rates or gustiness. The high frequencies, due to molecular causes, suffer a greater absorption in the open; on the other hand low frequencies are more damped in passage over the ground. When the atmosphere is not homogeneous in density at any level the variation is usually caused by clouds or fog banks, whereas an overall fog favours the perception of distant signals. This was brought out in the work done in 1870 by Tyndall for the Trinity House Brethren, using the siren at the South Foreland as source. Humidity is, in fact, an indicator of uniform atmospheric conditions rather than a prime cause of good propagation. It is "patchiness" of the atmosphere in the form of acoustic clouds, or even of large eddies, which makes for variation in audibility and, sometimes, a succession of echoes from a simple pulse.

Sometimes these patches, in the form of acoustic clouds, cause continuing echoes of the blast of a fog-horn lasting as long as 15 seconds after it has been turned off. Owing to the presence of such reflecting patches in the atmosphere a siren giving a continuous note of short duration is sometimes heard better and sometimes worse than the crack of a gun giving a short impulse.

Tucker, some years ago, explored the sound field round the North Goodwin and Casket sirens. During a typical summer day the intensity of the sound picked up varied irregularly in a 30 to 1 ratio over a period of a few minutes. In these cases, again, a reduction in humidity was generally accompanied by a reduction in audibility. A favourable temperature gradient for the propagation of sound was experienced when air had passed over a land surface at high temperature, and subsequently had its lower strata cooled by passage over the sea in the neighbourhood of the source.

When the atmosphere is not homogeneous in density at any level the variation is, of course, usually caused by clouds or fog banks.

Clouds or rain may also reduce the audibility (apart from molecular absorption) by the marked scattering which may ensue when the wave-length of the radiation becomes comparable with the scattering obstacles (cf. the corre-

sponding phenomenon in light). This, in fact, was the subject of a research, noted in a previous *Report*,* in which a beam of high-frequency sound was passed through artificial smoke suspensions. This work has been reviewed by Epstein (1941), who points out that the observed absorption is much greater than that predicted from molecular theory.

To return to the more recent work, Eagleson's results have been obtained in the open air at Bloomington, Indiana, U.S.A., over short distances (200 yards). Whereas theory and practice apparently agree within the surprising degree of one per cent in a uniform atmosphere, on a "bad" day variations in the intensity records ranged over 10 per cent of the normal, after allowance had been made for the effect of the average humidity on the anticipated absorption. Fluctuations of the same order were found by Tucker and by King in experiments made during the last war. Eagleson used a condenser microphone and preceded his open-air determinations by experiments in an absorbent box in which temperature and humidity could be independently controlled.

Sieg gives data which show that considerable fluctuations (60–80 db.) may occur even on a calm day in the sound received at a distance of 200 metres from the source. The effect of molecular absorption is shown in the following table of the fall in sound level to be expected over 100 metres at various frequencies:—

Frequency	250	500	1000	2000	4000 c. sec.
Transmission loss with molecular absorption	3·35	3·35	3·64	4·16	6·17 db.
Transmission loss excluding molecular absorption	3·35	3·35	3·19	3·20	3·28 db.

That low-frequency sounds are mostly damped by fricative influences was shown by placing a loud-speaker (emitting at 60 c./sec.) successively at 13 and 3 metres above the ground. A microphone at 100 metres showed a difference of 8 phons for the two positions when the experiment took place over a field of ripe corn, but only 3·5 phons difference after the corn had been cut.

Oswatitsch (1941) mentions another effect in connection with the transmission of low-frequency sound waves through clouds, namely, the possible influence of condensation on, and evaporation from, the drops due to their periodic contractions and expansions. If the frequency is high enough, the quantity of water which a drop can take up or lose in a half-period is too small to affect the commonly assumed adiabatic conditions, but at low frequencies some condensation of vapour and evaporation of water can occur during the compression and rarefaction portions, respectively, of each cycle. Moreover, when this does occur the condensation lags behind the expansion, so that density changes are no longer in phase with the pressure changes. There is, then, both absorption and dispersion of the velocity of sound because some of the acoustic energy is used up in the change of state. The velocity at high frequencies is then the

* *Rep. Prog. Phys.* **6**, 242 (1939).

same as that for dry air; that at low frequencies is the value corresponding to the wet adiabatic state. Actually, this interesting state of affairs does not occur except at infrasonic frequencies below 5 c./sec. In this connection we may note that Rocard (1940) had already suggested that the absorption of sound in the atmosphere might be due to work expended by the sound wave in evaporating water from small drops held in suspension in the atmosphere, but he found the work required to remove even a single molecule from a drop—electrically charged as most of them are—to be enormous compared with the energy of a sound wave. As Oswatitsch's analysis deals with uncharged drops, there may, therefore, be a further limitation, other than that of frequency, to the application of his reasoning to the absorption of sound in mists, and it seems more likely that the observed fall in sound intensity is due to scattering and the inverse square law rather than to the effect he discusses.

Bate (1940) describes a simple apparatus for the determination of the velocity of sound at low temperatures, consisting of a pair of bottle-type resonators. The necks are identical, but their lengths can be altered together by the same amount, while the cylindrical bottles always differ in length by 8·4 cm. It follows that when the two resonators separately excited by an air-jet have the same frequency, the half wave-length of the vibrations in the two resonators is 8·4 cm. Hence the velocity of sound for this wave-length can be calculated. As long as the temperature gradient along each neck is the same it does not matter if they jut out of the low-temperature enclosure. Using this apparatus, a value of 283 m./sec. for air at $-78{\cdot}5°$ c. is obtained (frequency 1685 c./sec.).

An interesting argument between Richter (1940), Schaefer (1940) and Bopp (1940) has been going on over the expression to be given to the pressure of radiation, when one includes terms of the second order, which are important when finite amplitudes are concerned.

Rayleigh envisaged a beam of sound, bounded at the sides by a tube and coming from and incident on discs normal to the axis of the tube (figure 1 *a*). Brillouin and Langevin assumed the sender and receiver to be facing each other

Figure 1 *a*.

Figure 1 *b*.

in an infinite medium, between which the beam passes without diffractions (figure 1 *b*). We further suppose that they are so spaced that stationary waves are set up.

The density of potential energy $e = \int \frac{dp}{\rho}$.

Expand p as a Taylor series:—

$$p = f(\rho_0) + f'(\rho_0) \cdot (\rho - \rho_0) + \tfrac{1}{2} f''(\rho_0) \cdot (\rho - \rho_0)^2 + \ldots\ldots$$

with $f(\rho_0) = p_0$.

The potential energy $E_p = \rho_0 \int_{\rho_0}^{\rho} \frac{dp}{\rho}$

$$= \frac{1}{2\rho_0} \cdot f'(\rho_0) \cdot (\rho - \rho_0)^2 + f'(\rho_0) \cdot (\rho - \rho_0) + \tfrac{1}{2} f''(\rho_0) \cdot (\rho - \rho_0)^2,$$

so that, approximately,

$$p - p_0 - E_p = (2\beta - 1) E_v - \rho_0 f''(\rho_0)(\rho - \rho_0). \qquad \ldots\ldots(1)$$

where

$$\beta = \frac{\rho_0 \cdot f''(\rho_0)}{2f'(\rho_0)} = \tfrac{1}{2}(\gamma - 1)$$

in the adiabatic state.

When we include the kinetic energy, E_k,

$$E_k - E_v + (p - p_0) = C, \text{ a constant.} \qquad \ldots\ldots(2)$$

In figure 1 *a*, since the tube cannot leak, $\overline{\rho - \rho_0} = 0$, and from (1) and (2) the mean pressure on the receiver,

$$\overline{p - p_0} = (\beta + 1)\overline{E}_p = \tfrac{1}{2}(\gamma + 1)E,$$

the receiver being at a node where $E_k = 0$.

The mean pressure on the side walls is $C = \beta E = \frac{1}{2}(\gamma - 1)E$.

In figure 1 *b*, where the sound field communicates with the undisturbed fluid, $C = 0$. So the mean pressure on the receiver $\overline{p - p_0} = \overline{E}_v = E$.

Thus, while in both cases under isothermal conditions the pressure on the disc is E, under adiabatic conditions it is still E in the Brillouin case, but $\frac{1}{2}(\gamma + 1)E$ in the Rayleigh case. In the practice of radiation pressure measurement the "unenclosed" type is more common, fortunately, as then the pressure is independent of the equation of state of the gas.

Also, in the practice of the measurement of radiation pressure, a constant has to be introduced to allow for conditions at the edge of the disc. This is a function of $2\pi a/\lambda$, where a is the radius, and has been formulated by King. Recently, Fox (1940) has tabulated this constant (for 5 relative values of density of solid and fluid) and for values of the above parameter ranging between 2 and 20. As a approaches $2\pi/\lambda$, the pressure on the obstacle diminishes, due to scattering of the radiation.

Unwin (1941) has worked on the theory of large displacements in relation to the waves produced by the sudden release of compressed air, acting on a suggestion of G. I. Taylor's that the effect is a combination of two occurrences: (1) the expansion of a solid sphere *ex nihilo* to a finite radius, and (2) the sudden release of a compressed sphere of fluid initially at rest. The solution of (1) is awaiting publication. The second is given under the assumptions of constant entropy and spherically symmetrical density, i.e., $\partial\rho/\partial r$ continuous from the centre to the uniform value which it has in the atmosphere outside. With such waves the finite amplitude falls off inversely as the distance and opposes the tendency of the wave front to steepen into the form characteristic of a shock wave. In fact, the assumed disturbance gives rise to a train of waves of diminishing amplitude rather like the ripples on a water surface when a pebble is dropped in.

Further, in the test case of the theory, the zones of compression are small in extent compared with the zones of rarefaction. When the initial disturbance reaches a certain intensity a shock wave will develop and the question, not yet treated, of whether it will remain such or degenerate to a normal wave form is raised.

Thompson (1940) enumerates the procedures to be followed in recording shock waves. Density is measured by spark photography or by an optical interferometer, displacement by a heated filament, pressure amplitude by the piezo-electric gauge, the Hopkinson bar or the condenser microphone. The first of these pressure instruments is probably the most accurate, and a means of calibrating it on a ballistic meter is described. Besides giving a résumé of the theory of shock waves, this author gives a table showing the magnitude of physical factors associated with typical shock waves, from which the following is extracted:

Source	Wave velocity	Condensation	Time of condensation	$= \frac{\text{Intensity press. amp.}}{\text{area}}$
Cal. 0·30 bullet	750 m./sec.	1·5	10^{-10} sec.	10^{6} watts/cm^2
Large projectile	750 ,,	1·5	10^{-9} ,,	10^{9} ,,
10 lb. T.N.T.	5000 ,,	2·8	10^{-12} ,,	10^{9} ,,
1000 lb. T.N.T.	5000 ,,	2·8	10^{-12} ,,	10^{8} ,,

§ 2. ULTRASONICS

(i) *Gases*

Herget (1940 a, 40 b) has designed an ultrasonic interferometer (figure 2) in which the quartz (C) and reflector (R) are kept at a fixed distance and in

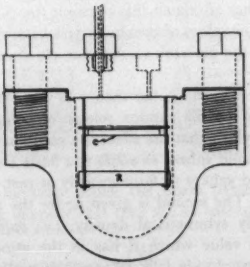


Figure 2.

which a gas may be put under pressure. This type of instrument possesses advantages over the more usual one in which the reflector is moved and the half wave-length estimated from the recurrence of peaks in the driving current,

for when gases under pressure are in question it eliminates the gland and the possibility of leaks through it, where the reflector guide enters the chamber. As the frequency of the oscillator is fixed as well as the path length, it is then necessary to alter the pressure until this path becomes equal to an integral number of half wave-lengths, as shown by the variations in the current through the driving circuit. The enclosure is built to withstand a pressure of 55 atmospheres. The value of the velocity of sound is supposed known at some pressure within the range at which the gas is first introduced. The gas is then allowed to leak out slowly through a tap and the number of peaks passed as the pressure is reduced is counted. To test the instrument the velocity in ethylene was measured both in it and in one in which the reflector moves, namely, the variable-path type of instrument. The two sets of data concur within one-half per cent.

Pielemeier, Saxton and Telfair (1940) have derived both the velocity and absorption coefficients in carbon dioxide, with which various percentages of water vapour were mixed, over the range of frequency 284 to 1595 kc./sec. As the humidity is increased, the absorption per wave-length rises to a maximum of about 0·28 and then diminishes. This occurs at each frequency. As the absorption peak is approached the velocity drops by about 10 m./sec. This is taken to indicate that the lifetime of a quantum of vibrational energy in the CO_2 molecule decreases with humidity. A useful comparison with allied phenomena is cited, as follows. The observed absorption at peak value is so great that a layer of carbon dioxide 12 wave-lengths thick absorbs 99 per cent of the ultrasonic energy. At 1595 kc./sec. (the highest frequency used) such a layer is just over 2 mm. thick. This is comparable with the best sound-absorbing materials for audible sound or of black glass for visible light, or of lead for x rays.

A number of investigators have dealt with the effect of temperature on ultrasonic velocity. Bender (1940) used a single quartz operating each way against a pair of reflectors, one each side of the quartz. If a is an apparatus constant to represent errors due to spreading of the radiation etc., and if one reflector is set at a distance d_1, corresponding to n_1 half wave-lengths and the other at d_2, corresponding to n_2 half wave-lengths, we have $d_1 = a + n_1\lambda/2$ and $d_2 = a + n_2\lambda/2$, so that the distance $d_1 - d_2$ is equal to $(n_1 - n_2)$ half wave-lengths.

Bender then kept the reflectors fixed and raised the temperature until another integral number of (new) half wave-lengths intervened. In this way he got values (V) for velocity in nitrogen, carbon monoxide and nitric oxide at 1000 kc./sec. On plotting $V/T^{\frac{1}{2}}$ against T (temperature) he obtained a curve which rose to a maximum and then fell. This dispersion region should move to higher temperatures as the frequency is raised, according to Bender, though actually he used only the one frequency in this research.

Overbeck and Kendall (1941) made measurements in pure carbon dioxide gas at temperatures ranging from 25° to 530° c. and frequencies from 27 to 147 kc./sec. These authors confirm that the velocity (adjusted to 0° c.) falls as temperature increases, but, contrary to results by Penman and by Warner, their work suggests a much more gradual fall. The difference is ascribed to impurities

and temperature gradients in the earlier work. It should be pointed out that their room-temperature value of the reduced velocity is considerably above the accepted audible value, though their frequency was only 90 kc./sec.

Leonard (1940) describes a method for measuring the absorption in a gas based on the system of Grossmann, in which a magneto-strictive source operates in a baffle and the intensity along the axis of the source is recorded by means of a microphone. The whole is mounted in a cylindrical chamber, of which the walls are covered with baffles to prevent reflected energy reaching the detector. Until this improvement was introduced the distance-intensity curves as the microphone was withdrawn showed "corrugations" The microphone was a "bender" built up from a bimorph element of Rochelle salt having a natural frequency of 45 kc./sec. To prevent reflections from its surface interfering with the source, it had a surface of only 3/8 in. square. Work has been begun on carbon dioxide, nitrous oxide, carbon disulphide, carbonyl sulphide and sulphur dioxide. For the linear molecules a linear relationship was found between the maximal absorption coefficients and the frequencies at which they occur. Great care was taken in the purification of the gases. In the same laboratory, Fricke and Knudsen (1940) have found how the absorption of these gases is affected by "acoustic catalysts" in the shape of molecules of hydrogen, water, sulphuretted hydrogen or methyl alcohol. The peak of absorption is thereby shifted to higher frequencies by an amount proportional to the concentration of the adulterant. While Eucken and Aybar (1940) have used methane and carbonyl sulphide in mixtures, van Itterbeek and Mariens (1940) have worked in the same direction, tracing the effect of ordinary or heavy water as catalysts to the absorption in oxygen and carbon dioxide.

The method of Petersen (1940), developed further by its originator, of studying the propagation of ultrasonics in gases, using the diffraction of a monochromatic light beam, has been applied by Keller (1940) to a determination of the absorption coefficients of argon, nitrogen, ammonia and carbon dioxide and their variation with pressure at a frequency of 4 Mc./sec. With the exception of argon, which conforms to theoretical prediction, the absorption is high compared with what would have been expected. Petersen himself discusses the accuracy of measuring the ultrasonic intensity from the brightness of the diffraction pattern by photometry. He prefers to turn the axis of the sound beam slightly, relative to the optic axis, so that the angle between them is progressively lessened, taking photographs of the optical beam the while. Some inconsistencies in gases between intensities at different distances from the source are to be ascribed to the "electric wind" set up by a strongly vibrating source in a gas. This is a repulsion of the gas molecules by the quartz which tends to disturb the sound field.

At this point we may include some work by Kneser (1940) on the absorption in nitric oxide, in spite of the fact that it was done with an audible source. He used Oberst's method, in which the resonance curve of a cylindrical resonator (a brass tube 4·5 cm. radius and 56 cm. long) in a thermostat is derived from the behaviour of a loud-speaker source at one end and a condenser microphone at

the other. The spread of the resonance curve (and therefore the damping) was found to vary inversely as the square root of the absolute temperature—as would be expected on classical theory. Kneser was looking for deviations from this due to electronic relaxation times. Although nothing of this sort showed up, within the limits of experimental error, it was possible to deduce an accurate value of the velocity in nitric oxide (in the range 300 to 5000 c./sec.).

(ii) *Liquids*

Dealing first with optical diffraction methods, Bär (1940) applies the Jamin interferometer for the measurement of the diffraction bands caused by the passage of the light athwart the ultrasonic beam. He points out the necessity for ensuring that only coherent light rays interfere, and also the need for large sound intensity in such work if the absorption coefficient is to be accurately obtained, though this may mean that the coefficient so obtained does not truly refer to small ultrasonic amplitudes. Gregg (1941) used the Lucas and Biquard method, but instead of measuring the shift of diffraction bands when the ultrasonic source was turned on, he changed the driving voltage on the quartz oscillator until the light was brought back into focus in the viewing microscope. The light rays as they pass through the ultrasonic beam have, in fact, their radius of curvature changed. A linear relationship emerged between the logarithm of the voltage that had to be applied to re-focus and the distance down the ultrasonic beam from the source. The absorption coefficient was measured as a function of the slope of this line.

A new method which does not demand large source outputs has been developed by Richardson (1940 a). In this a stationary wave is set up between a quartz oscillator and a movable plane reflector and the interspace explored by traversing a thin heated platinum wire along. From the position and velocity amplitude at the antinodes both propagation constants may be derived. The hot wire was calibrated for response by varying the output of the source. Values of velocity and absorption in water, hexane, carbon disulphide and benzene at frequencies between 200 and 1100 kc./sec. are quoted.

Narasimhaiya and Daraiswami (1940) have adapted the optical diffraction method to measure the velocity in liquids by using white light and a spectrometer. The spectrum of the iron arc is superposed and used for calibration. The diffraction due to the passage of ultrasonics through the liquid is made perpendicular to this and gives rise to lines which spread out towards the red end. The wave-length of the ultrasonic beam in the liquid can be found at any point where the diffracted lines are cut by a recognized iron line. Krishnan (1940), and also Bhimasenachar and Venkateswarlu (1940), have measurements in about 20 organic liquids each. The former finds no dispersion of velocity between 3000 and 8000 kc./sec., and the latter note that the velocity is usually higher in the aromatic than the aliphatic esters. Sibaiya and Narasimhaiya (1941) have measurements in solutions of naphthalene and Suryaprakasam (1940) of the temperature coefficient in 11 organic liquids. c falls by 3 to 5 m./sec. for each 1° c. rise of temperature.

Prevailing uncertainty here is brought out by Schaafs' (1941) criticism of Parthasarathy's results, which he maintains do not show that aromatic substances have a generally higher velocity than aliphatic. Increasing density through a homologous series ought to decrease c, but, in fact, does not. In halogen compounds the abnormal increase in molecular weight accounts for the observed fall in c through a series. Increase in molecular volume and the density go hand in hand. These are the two factors which cause the increase of c when heavier atoms are substituted in esters, and not increase in molecular weight. There is no real evidence that dipole moment is of importance for rise of c.

Rao and Ramaiya (1941) remark that on the Herzfeld theory the enhancement of heat conduction at high frequencies should cause a fall in velocity away from the Laplacian (adiabatic) towards the Newtonian (isothermal). The critical frequency for this drop should be given by $c^2\rho s/4\pi k$ (s = spec. heat, k = conductivity). Substitution gives the figures of 10^{12} c./sec. in water and $5 . 10^8$ c./sec. in a number of gases for this critical frequency.

In practice there is a drop of velocity, which the authors ascribe to this cause, at about 10^9 c./sec. in certain liquids and at 10^6 c./sec. in gases like carbon dioxide and nitrous oxide. Herzfeld (1941) points out that because of the difficulty in measuring small changes in ultrasonic wave-length it might be better in liquids to "spot" any such change from a change in the parameter absorption/frequency^2, but even so, it is doubtful, under present limitations of technique, whether one could detect a change of 10 per cent in this parameter.

Willard (1941) gives values of velocity and absorption in forty liquids, based on optical diffraction methods, in the range 6 to 30 Mc./sec.

Besides these we find absorption measurements on water between 7 and 50 Mc./sec. by Fox and Rock (1941) using the deflection of a bead suspended at various distances from the source, which had, of course, to be "run" at high intensity, so that the "electric wind" to which we have already referred troubled the investigators until cut off by a thin collodion film. Beam divergence and multiple reflections from the container were also troublesome. Some high absorption values were ascribed to cavitation. (This paper is worth studying since it gives a rather more candid exposition of the difficulties and sources of error latent in the technique of absorption measurement at high frequencies than one usually finds [In the opinion of the author of this *Report* all measurements of the velocity and absorption requiring high intensity at the source are open to suspicion.] The unpalatable truth is that measurements in the ultrasonic field, at present, have not the precision of spectrometry, and little is to be gained for the science of acoustics by glossing over this fact.)

Hunter (1941) uses the reaction of the liquid column upon the circuit driving the quartz oscillator to measure the absorption in glycerine and several viscous oils. He also animadverts on the discrepancies between results obtained by different methods and in different laboratories.

Hartmann and Focke (1940) refer to Sewell's expression for the absorption of sound in a viscous atmosphere by small spherical obstacles, whose circumference

is small compared with the wave-length and whose volume concentration is small. The expression for the intensity coefficient of absorption is

$$n\pi a^2\left[3\sqrt{2}\frac{\sqrt{\nu\omega}}{c}+\frac{6\nu}{ca}+\frac{7}{9}\left(\frac{\omega a}{c}\right)^4\right],$$

where a is the radius of the obstacles, n their concentration, c the velocity of sound, ω the pulsatance and ν the kinematic viscosity of the fluid. The first term represents the loss due to turbulence, the second due to skin-friction and the third scattering. In ultrasonics the motion of the obstacle may usually be ignored (though this is dealt with in an appendix to Sewell's theory). The experiments involved intensity measurements by deflection radiometer. Aqueous suspensions of lycopodium spores were used. Measurements in pure water indicate an absorption band between 194 and 3000 kc./sec. The turbulence term in the theory is the most important, but the coefficients of absorption as measured were only about half the predicted value. Epstein (1941) also refers to this question, pointing out the large discrepancy between Laidler and Richardson's results for aerosols and the values deduced from Sewell's theory.

(iii) *Miscellaneous ultrasonic effects*

Oka (1940) works out the effect of an ultrasonic field on smoke particles of stick shape and shows how their orientation can produce double refraction of a beam of polarized light. The phenomenon has been observed by Hawamura in a vanadium pentoxide sol. The anisotropy is due to the tendency of the particles to orientate themselves in the field. This effect quickly vanishes after removal of the field, since such small particles resume their promiscuous orientations after a short relaxation time, unless they are submerged in a viscous medium. (The present paper is purely theoretical.)

Shear and Focke (1940) have used ultrasonics to excite a rod into longitudinal vibration (as in Kundt's tube), and by means of lycopodium powder found the resonant wave-lengths for various frequencies from 120 to 750 kc./sec. One end of the rod is supported by a silk thread, the other passes into an oil bath into which the quartz oscillator is working. In this way they have measured velocities in silver, nickel and magnesium.

Seidl (1940) showed that the electrical conductivity of liquids like hexane, toluene and carbon tetrachloride is changed when irradiated with supersonics. He immersed two parallel-plate electrodes vertically in the liquid and set the beam up between them. A difference of potential of 20 volts was put upon the plates and the leakage measured by a quadrant electrometer. Adequate control had to be kept on temperature, but when this was allowed for, the radiation caused an increase in the leakage current in every case. The effect is attributed to ionic combination and breaking up of molecular complexes.

§3. VIBRATING SYSTEMS: SOLIDS

(i) *Bowed strings*

In a paper with the intriguing title: "A scientific search for the secret of Stradivarius", Saunders (1940) describes five years of work "to find out how

violins of unquestioned reputation act when played in the usual manner and to try to deduce from these how the violins vibrate". The hope is expressed that this work will help in supplying the world with adequate substitutes for the greatest instruments when these wear out during the next two centuries. About a dozen of the very best old instruments were compared with thirty good and bad modern instruments. It was found at the outset that the antiques had required strengthening and sometimes re-conditioning during their history to enable the increased tension of the strings consequent on the progressive rise of pitch which has taken place to be borne, and also to enhance the intensity which modern practice requires. The quality of a violin differs from note to note. In order to establish a standard bowing technique, the author required the emission of each note at a steady intensity for five seconds (one bow length). This was accomplished by a human player having such a steady arm that he could bow without noticeable variation in quality during the stroke, so that the personal equation of the artist was held to have been eliminated in the trials. The usual procedure was to analyse, on a specially built harmonic analyser, each note twice in an octave of semitones on each string. The response curves so obtained differ because of variations in dimensions and wood. The best have a fairly uniform response level over most of the range, with a little excess near 2 kc./sec. and a drop at 10 kc./sec. By using a weighted mechanical player (in the form of a rotating celluloid disc) it was found that the violins characterized as excellent required less pressure to bring out adequate intensity at good tone quality, the old being easier to play than the new in this respect. The difference in quality *per se* between good modern and old violins is indistinguishable. This was shown by a test in which three violins (one a Stradivarius) were played behind a screen. One-third of the audience correctly picked the Strad, which showed that the voting was by pure chance.

Rohloff (1940) has also obtained acoustic spectra of violins. In addition he has measured the rate of damping of notes produced in air and *in vacuo*. If the air is responsible for a high proportion of the total damping and the internal friction small, the apparatus is an efficient radiator of sound. This is the case with the old Italian violins examined. By comparing the decrement of beams, stressed and unstressed, in transverse vibration, this author concludes that the makers stressed the bellies of their instruments to produce this effect—but compare the remarks of Saunders, above, on this point. The effect of the varnish was to reduce the internal damping in the wood.

In order to amplify the sound of a violin Vermeulen (1940) fixes a pick-up just above the bridge. The magnet supplying the field for this coil is secured by cords to the ends of the instrument. His object is to replace the ' choral effect" of a number of fiddles playing the same part by a solo instrument amplified to give the same intensity. Whether this is musically desirable is a moot point.

(ii) *Struck strings*

Having found by means of a pick-up probe that a soundboard in uniform compression gave the best response, Bilhuber and Johnson (1940) devised an

instrument for measuring the vertical component in the tension of the strings of a pianoforte, and were then able to make adjustments to the compression to make this an optimum. They also made experiments designed to reduce the mass and stiffness of the soundboard by tapering it. The new form gave a brilliance to the upper treble by increasing the average duration of tone after a standard key-striking impulse by 30 per cent. Further, the overall intensity was increased.

(iii) *Plates*

A number of investigations on Chladni figures fall within the period under review. Mary Waller (1940) has continued her work on free square plates by noting the compound modes produced when two sets of nodal lines are established simultaneously. She points out that irregularities in the figures may be produced by constraints at the central support or by variations in the properties of the material over the plate. Since the theory only applies when the thickness is small compared to the distance between adjacent nodal lines, one expects discrepancies between theory and practice in the higher modes, though these degraded modes are reached sooner with the circular than with the square plate. Colwell, Stewart and Arnott (1940) show how to calculate some of these compound modes and compare the mathematical figures (for the circular plate) with some of Chladni's original experimental drawings.

(iv) *Reeds*

McGinnis and Gallagher (1941) set up an apparatus for blowing a clarinet reed artificially. The box in which the reed was housed was equipped with glass windows through which the reed could be observed. It was found, contrary to what has been assumed in most theoretical writings on the instrument, that the embouchure was completely closed by the reed for a large part of each cycle, so that the current of air was cut off. At normal blowing pressure the reed was only in motion for about a quarter of a period, the rest of its time being spent at the two extreme positions. Zahradnicek and Nespor (1940) noted another feature of a beating reed, i.e., that it has a negative characteristic in the sense that if one plots blowing pressure (in the wind-chest) against the current of air past the reed, the curve has a negative slope in the same way that potential and current are related in the electric arc.

§4. VIBRATING SYSTEMS: FLUIDS

Hilton (1940) has made measurements of the noise from aerofoils and streamline wires exposed to the draught of the high-speed wind tunnel at the National Physical Laboratory, which, not having an actuating airscrew, but being driven by compressed air, is eminently suitable for such work. The noise of the unimpeded airstream in this tunnel consists of the hiss of the jets which, when the tunnel is being run at high speeds, may not reach the observer, as it is carried down the tunnel by the shock wave which develops when the speed of sound is approached locally.

The sounds of the obstacles were of two types:

(i) A sound of definite frequency, ascribed to the production of aeolian tones, very near the upper limit of audibility at the high speeds involved. They were often accompanied by unpleasant beats due to variations in conditions along the wing span.

(ii) A loud steady noise (resembling that of a factory hooter) which, on thick aerofoil sections, sets in at a speed of about 800 ft./sec. It was estimated to have a level of about 120 phons and to be between 200 and 300 c./sec.

Both noises were reduced by roughening the surface (evidently by turning the aeolian vortex street into an incoherent turbulence), though the second source of noise could only be eliminated by reducing the ratio of thickness to chord of the strut below 15 per cent. This noise appears to be associated with the formation of a shock wave, since the drag coefficient of the body suffers its characteristic large increase at the same speed.

Lenihan and Richardson (1940) have experimented on a phenomenon which combines both aeolian and edge tones, i.e., the sound which ensues when a jet of air leaves a narrow slit to fall upon a thin wire which is set parallel to it. The sound apparently comes from the interaction between the Bénard-Kármán avenue of vortices which start out from the slit and another set which march from the wire itself. When these fall into step a pure tone is produced. Increasing the wind speed results in jumps to the octave, etc., as with the common type of edge tone. If the wire is set somewhat to one side of the jet, where there is a transverse gradient of velocity, it can itself be set in vibration. Savic (1941) points out that an important difference between the simple aeolian and edge tone phenomena is to be sought in the variation of velocity in the direction of the stream, which is more significant in the latter case. He derives the stream function for a linear jet (including viscosity) and shows that it causes the stream to pendulate in a manner which agrees with Brown's photographs of jets and also with measurements of the " wave-length " of the vortex system associated with edge tones. A free jet is then sensitive to sound because its velocity profile has a point of inflection; this profile can be modulated by an external source of sound.

In his Guthrie Lecture, Andrade (1941), after reviewing the history of the subject, describes experiments with sensitive jets, both gaseous and liquid. A loud-speaker was used as source of sound. When the gas is ignited, as in a Bunsen flame, it is the inner cone of unburnt gas which is sensitive, the flame itself being merely an indicator of the former's movements. If this cone is absent, as in a hydrogen flame, the system is insensitive. Fundamental for the understanding of the phenomenon is an experiment showing that a sound travelling along the axis of the flame, actually from a loud-speaker above, produces no disturbances, whilst the same sound travelling transversely produces typical flaring. This and other experiments show that the disturbance of the sensitive flame is produced by a transverse periodic relative motion of the medium and the nozzle.

In experiments in flames or water jets it is necessary to eliminate carefully resonances in the structure of the apparatus, as otherwise the response frequency curve of the jet exhibits sharp peaks, which may also be produced by resonant frequencies of the room itself. This was shown by subjecting a sensitive jet of water into water to disturbances of all frequencies from 1·4 to 1165 cycles per second, the lower frequencies being produced by a motor with an unbalanced flywheel. Within this range 54 resonances were found, and the resonance frequencies were not shifted by changing the liquid or the diameter of the orifice. On the other hand, when the jet itself was vibrated by attachment to the diaphragm of a loud-speaker, a smooth response resulted.

To study the theoretical behaviour of a flame jet, as represented by two parallel vortex sheets, use was made of an electromagnetic analogy, the magnetic field round a line current being equivalent to the velocity field round a line vortex. A system of vertical wires was arranged so as to represent a two-dimensional vortex street, each side of the street being equivalent to a vortex sheet if the wires are fairly close. An electric current passed up those on one side and down those on the other. An exploration of the magnetic field in a plane perpendicular to the wires gave, by analogy, the corresponding flow. In this way the velocity at any point was obtained for a given distribution of rods, and to obtain the distribution corresponding to a later instant the position of each rod was displaced by an amount, and in a direction, corresponding to the velocity deduced. This procedure showed that the simple theoretical two-dimensional jet ultimately developed spurs and curls, and, in fact, similar spurs and curls were produced experimentally by a flat jet of water into water, when transverse vibrations of the flat nozzle were set up.

A flat jet is, however, not nearly as sensitive as a round jet; further, a jet from an elliptical orifice has greater sensitivity when the sound is coming along the direction of the minor axis than when it is coming along the major axis. This is attributed to the fact, proved theoretically, that in the round jet the vorticity diminishes inversely as the square of the distance, but in the flat jet inversely as the distance from the effective source of the jet within the nozzle. The disturbance of the jet due to sound, or any relative periodic movement of nozzle and medium, develops from the point where the jet leaves the nozzle: the disturbance due to excess velocity alone develops from the end of the jet remote from the nozzle. Favourable conditions for sensitiveness are, then, heavy vorticity in the neighbourhood of the nozzle and small vorticity at the remote end of the jet.

In general, it has been shown by Andrade that, if the jet is treated as a vortex sheet, subject to transverse disturbance when it leaves the nozzle, the familiar instability can be explained on the basis of classical hydrodynamics, without any of the special assumptions hitherto made.

Hartmann and Lazarus (1940, 41) have been making measurements of the radiation from the well-known air-jet generator devised by the former. This emits high-frequency sound from the collapse of shock waves into a resonant cavity in a jet moving at speeds greater than sound. The amplitude of the aerial

vibrations within the cavity are made visible in virtue of a phenomenon similar to that noticed by Hilton, i.e., that the large pressure drop through the front of the shock wave causes the wave-front to be visible. In fact, it acts as a "Riemann mirror". Photographs are given which show this as a bright line between the orifice and the resonator-oscillator. Oscilloscopic records of this bright line show it to be oscillating to and fro with the frequency of the tone emitted. To the unaided eye this line is, in fact, thickened to a degree which indicates the amplitude of the aerial vibration in the system. This amplitude is compared with the intensity in the neighbourhood of the generator as given by the deflections of a Rayleigh disc.

Schüster and Matz (1940) revert to the circulations in Kundt's tube, the study of which was a fruitful subject of research a decade ago. They derive the following equations for the radial (u) and axial (w) component of streaming velocity at a point specified by cylindrical co-ordinates r and z:

$$u = \frac{3}{8}\frac{kr}{c}\frac{P_0^2}{c^2\rho^2}\left(1-\frac{r^2}{R^2}\right)\cos 2kz,$$

$$w = \frac{3}{8}\frac{1}{c}\frac{P_0^2}{c^2\rho^2}\left(1-\frac{2r^2}{R^2}\right)\sin 2kz$$

(R is the radius of the tube, P_0 the nodal pressure-amplitude, $k=\omega/c$).

These equations specify the familiar characteristic of flow from node to antinode near the axis and back along the wall, w being zero at $r=0{\cdot}707\,R$. The maximum current is on the axis midway between a node and an antinode, being $w_0=3/8\,.\,U_0^2/c$, where U_0 is the antinodal velocity-amplitude in the S.H.M. Using cigarette smoke, values of w_0 for different values of P_0 were obtained (the latter by a manometric membrane) and found to give good agreement with theory, such differences as exist being explained by the smoke particles, owing to their finite size, not having the full amplitude of the sound waves.

Oberst (1940) sets out to produce high intensity sounds in a column of air, using a siren source at one end on the same principle as the diaphone used in fog signals. In order to get pressure amplification, a transducer in the form of a reduction of the cross-section is introduced at one point in the tube. In the simple case when the length of each tube is a quarter wave-length long for the radiation in question, then at the end of each tube a pressure amplification equal to the reciprocal of the length of tube times the damping coefficient is introduced. Thus one can, with the help of the sudden change of cross-section, produce a stationary wave system in the narrow tube with a very large velocity amplitude at the point where it joins the wide tube and a corresponding pressure amplitude at the closed end of the narrow tube. The pressure amplitude at the latter point was measured by a U-tube manometer, in which a steady pressure was built up. To prevent the d.c. air from the siren getting into the narrow tube, a rubber membrane was placed at the junction. On one system which he tried, one tube was 107 cm. long and 12 cm. diameter, the other 45 cm. long and 2·5 cm. wide. At a resonance of 186 c./sec. the pressure amplitude at the terminal node was 9×10^4 microbar, corresponding to a particle amplitude at the antinode of

2 cm. Various applications are suggested: the tracks of smoke particles within the tube could be photographed; the system could be used to calibrate microphones for larger amplitudes.

Ganitta (1940) tackles a rather similar problem: the transmission of sound through a column of water. First he used tubes of rubber about 10 m. long and 5 or 10 cm. wide with an electrodynamic hydro-sounder at one end and a crystal microphone at the other. At 300 c./sec. there was a fall of 10 db. every two metres in the latter tube, so that the waves showed a progressive system without appreciable reflection (damping coefficient about 10^{-3} cm.$^{-1}$). Later glass tubes 1 m. long were used. These showed stationary waves (at the appropriate frequencies) as the damping was only 4×10^{-8} cm.$^{-1}$. The air-content of the water was found to have considerable effect on both velocity and damping. Evidence of the existence of a wave-system in the glass as well as in the water was obtained.

Kastner (1941) has also studied this matter from the point of view of the application of the Helmholtz-Kirchhoff "tube correction" to low-frequency pressure waves in pipes. The waves were produced by a reciprocating piston having a stroke of 4·7 in. in a cylinder of 5 in. bore. Pipes having diameters from 1 to 3 in. were attached. Pressures at a distance were measured by a valve which opened to a mercury manometer at the same phase of each stroke. The velocity of propagation was independent of frequency between 20 and 150 c./sec., but the reduction in velocity with pipe diameter takes place more rapidly than for amplitudes to which classical theory applies, such as those used in the experiments of Fay (1940) (up to 5000 c./sec.) with carefully calibrated telephone source and terminal microphone.

The extent to which the walls confining a vibrating column of air may take part in the vibration has been studied by Knauss and Yeager (1941), using a cornet having a crystal pick-up attached to it, but though metallic vibration was detectable it proved to be insignificant as a contribution to the total sound output of the instrument. Boner and Newman (1940) have tackled the same problem from the point of view of organ (flue) pipes. The materials used ranged from steel to shellaced paper. Though some influence on overtone content was marked, the general effect on output was smaller than that reported by D. C. Miller.*

Tyte (1940) refers to an earlier paper in which he has calculated the end-correction for a conically flanged cylindrical pipe by analogy with corresponding problems in a.c. technology. He now shows that the experimental results of Bate and Wilson fit this theory well and has extended the experiments to deal with cones of smaller apical angle, for which the cone is in the nature of an extension of the cylinder rather than a flange to it. On the basis of this agreement he goes on to treat the classic case of the orifice in a diaphragm forming the open end of a pipe, and suggests that the end-correction may be taken as inversely

* His death on 21 February 1941 removed a fine personality from the confraternity of acoustic investigators. His pioneer work on the acoustics of musical instruments and the voice will always be quoted.

proportional to the area of the orifice, and tries out this idea on the experiments of Irons and Bate and Robinson. The suggested relation between end-correction and reciprocal area holds whether the pipes are flanged or not.

Taber Jones (1941) is also working in this field. He compares a number of equations which have been put forward with experimental facts. Most of these fit particular cases—e.g. holes in the side, flanged ends, and ranges of frequency—but none is universally applicable. While realizing the difficulties in applying an end-correction, mainly due to the uncertainty in the true value to be assigned to the physical constants involved in the expression, the writer agrees with this author that it is a concept worth retaining.

There are a number of accounts of "spectrum analyses" of musical instruments. Northrup (1941) has studied the quality of an open organ pipe in a "dead" room with a microphone at various positions in relation to the source. The results show that there is no single position at which representative spectrograms can be made. He thinks it may be possible to rotate the pipe in a very "live" room while readings at a fixed microphone are taken. This research is, of course, very important for the correct acoustic assessment of the quality and output of musical instruments, and it is to be hoped that it will be continued

§5. ACOUSTIC IMPEDANCE: PROPERTIES OF MATERIALS

(i) *Filters*

Lindsay (1941) continues his theoretical work on the transmission of compressional waves through a stratified medium, by considering that the transition in properties from one layer to the next is a gradual instead of a sudden one.

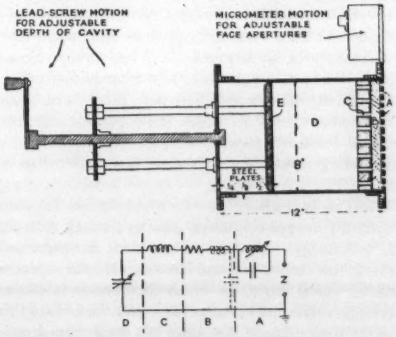


Figure 3.

It results from this assumption that the system loses its filtration characteristics. At least a definite gradient in elastic properties along the line is necessary if attenuation is to be a function of frequency. Labaw (1940) has a theory for the

passage of sound through parallel conduits, terminated by conventional impedances such as cylinders and cones. The power transmission for these, as a function of frequency, is shown in graphs.

Two papers show how to construct variable boundary impedances for this type of work. The impedance described by Bolt and Brown (1940) is on a large scale and intended to form part of a boundary wall. The mass element is provided by boards perforated with small channels of variable area, while stiffness comes from cavities of fixed or variable depth in a honeycomb behind and resistance from layers of porous material. Figure 3 shows the scheme (in section) with the appropriate electric circuit. A behaves as a variable inductance of magnitude $\rho l/S$ (l and S being the effective length and cross-sectional area, respectively), B as a series resistance, C as additional inductance and D as a capacitance $d/\rho c^2$ (d, the depth of the cavity, is less than a quarter wavelength). The impedance of the back plate E is not included in the discussion. The system is calibrated for different adjustments of S and d. E. C. Jordan's (1941) variable impedance is intended as a terminal to a pipe (figure 4). The

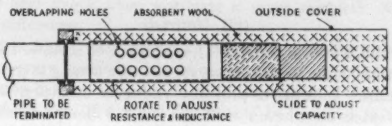


Figure 4.

resistance and inductance are simultaneously adjusted by sliding one set of holes over another to decrease the areas of their respective openings (like those in a pepper sprinkler). In order to overcome the change of inductance this has to be "tuned out" with an equal and opposite capacitive reactance (performed by a piston working in the terminal cavity).

(ii) *Impedance calculation and measurement*

Beranek (1941) adopts the tube method as a precision apparatus for the measurement of impedance, using at one end a loud-speaker (point) source and crystal microphone nearby. Unlike Hunt, who first used this method, he keeps the frequency fixed and varies the length of the tube (by means of a piston) and obtains the resonance curve as a function of the pressure amplitude (at the fixed microphone) and the length. The impedance is determined from the width of this curve at a given amount of anti-resonance. Corrections are introduced for (*a*) "tube" attenuation—Helmholtz-Kirchhoff correction, (*b*) reaction on source. He exhibits values of the specific normal impedance for a number of common materials. Chandra Kanta (1940, 41) uses the reaction on the loud-speaker source, measuring its electrical impedance under different conditions with the same object of measuring the acoustic impedance of the nearby material. Results for "Treetex" and felt are exhibited as a function of frequency. As a

check on the method, the impedance of a Helmholtz resonator was also measured, and the relations between phase and wave-length on the one hand and reactance on the other exhibited as curves. She also made up an "ideal absorbent" consisting of a cluster of narrow glass tubes let into the terminating piston. The absorption coefficient (per unit area) increased with the number of tubes, but there was no appreciable increase in the phase change at the reflecting piston with number of tubes (1 mm. diameter).

The attenuation in a single capillary tube is known from the theories of Helmholtz and Kirchhoff. The parameter which expresses the depth to which viscous and heat conduction losses degrade the sound energy is $(4\pi\nu/f)^{\frac{1}{2}}$, this being the thickness of the viscous "boundary layer" in the gas where it marches with the solid wall of the tube (ν = kinematic viscosity, f = frequency). At a sufficiently high frequency in narrow tubes this boundary layer extends over the whole bore of the tube, and the energy is so rapidly degraded that nothing returns from it into the main tube. At lower frequencies the tube is a reactive impedance, and there is both a loss of amplitude and a change of phase at reflection, dependent on frequency.

(iii) *Materials*

Rschevkin and Terossipjantz (1941), as a result of acoustic treatment of the dome in the Palace of the Soviets, advocate fine wire netting to produce frictional damping on walls. In order to provide a resistance coefficient independent of frequency, the spaces must have a width less than 0·3 mm. (if circular) or 0·2 mm. (if slits). If the grid has a depth of 1 mm. it is possible to get a screen having a resistance of 10 acoustic ohms per sq. cm. Actually, the reactive component of impedance imposes a greater restriction on the size of the net, since it must not exceed the resistance; this requires a thickness less than 0·13 mm. if the impedance at 4000 c./sec. is not to be more than 3/2 that at low frequencies. They suggest using a very thin foil (0·05 mm. thick) punched with holes 0·2 to 0·3 cm. diameter to attain this object. They did some tests on these, using the tube method, but in checking with theory it is difficult to estimate the end-corrections of the holes. They think one must take the effective length of the perforations to be about $2\frac{1}{2}$ times the thickness of the foil.

Erwin Meyer (1940) and his colleagues advocate a wall absorbent in which the "pores" are tapered channels. This is secured by covering the walls with large pyramids of rock wool, having their vertices facing outwards. In this way an effective "dead" room was constructed and tested for its 'free-field" characteristics by showing that the sound level decreased inversely as the distance from the source, as for spherical radiation, except at the lowest frequencies.

Kaye and Evans (1940) report measurements of absorption coefficients in the National Physical Laboratory reverberation chamber of materials such as turf, snow, gravel and ashes commonly found out of doors. In addition they have done some tests of existing and possible ballast materials for railway tracks. From the point of view of noise reduction, ashes would be an improvement as a stratum laid on the granite chips in current use, though, as the authors point out,

the road would have to be constantly watered to prevent the surface being carried away by the natural wind or the draught set up by trains. It would even be better to let the grass grow in the permanent-way if the foreman ganger could be persuaded to overlook the untidy appearance of the track!

A number of writers remark on the variation of absorption coefficient with area. Kawashima (1941 a) points out that the coefficient for an open window depends on the relation of its area to the volume of the room. Morse, Bolt and Brown (1940), also Sabine (1941), show the relationship between absorption and acoustic impedance. The satisfactory agreement between the solution of the equations of wave motion within the material and recently measured absorption values suggest that flow resistance and porosity are constant factors from 200 to 6000 c./sec. However, the train of deduction leading from these physical factors of the material to its absorption coefficient is such a long one as to make it impossible to reverse the process and deduce the factors themselves from measurements of the absorption coefficient. One must distinguish between normal coefficient, which is operative in the early stages of reverberation due to standing waves between opposite walls, and grazing coefficient, smaller and relevant to the later stages, when scattering is more complete. Cremer (1940) also notes this difference.

Parkinson and Jack (1941) discuss an important question for the acoustical engineer. Can every specimen be characterized, for practical purposes, by quotations relating to a few frequencies, and is an average of these of any value? It appears to them that, if a single figure is to be assigned to an absorption coefficient, that at 4096 c./sec. should be chosen, though it should be remarked that at frequencies in the neighbourhood of 4000 c./sec. the air plays an important part in the decay of sound in the room.

§6. SUBJECTIVE ASPECTS OF SOUND

(i) *Speech*

Both Bartholemew (1940) and Richardson (1940 b) discuss the mechanism of human speech, in particular the reinforcement available from cavity resonators, in relation to elocution and singing, the former being interested in the teaching aspects. The latter, and also Lewis and Tuthill (1940), deal with the interpretation of sustained vowel oscillograms produced at different pitches. In voicing "Oh" and "Ah", two low-frequency resonances (near 600 c./sec.) and at least two high-frequency resonances (near 2800 c./sec.) are typically employed. The high-frequency resonators are unique for each singer. Fairbanks (1940) has investigated a number of pitch-variation phenomena in voices which are interesting from psychological and physiological points of view, but are not so analysed as to reveal any fundamental physical characteristics. Obata and Kobayashi (1940) have applied their direct-reading pitch and intensity recorder to gramophone records of chanson and jazz singers, to observe the extent of the vibrato employed by these two types, and to the accuracy of intonation in the recordings of famous sopranos. They confirm the general impression among

those interested in the subject, i.e. "that all artists sing in a musical scale which corresponds neither to the equi-tempered nor to the enharmonic (or just) scale, but in a scale that is characteristic of the particular singer".

Drew and Kellogg (1940) determined the rates of rise of amplitude in speech sounds, which are, of course, associated with the voiced consonants preceding a vowel. Curiously enough, it is not the explosive vowels, p, t, k, which give the most abrupt start. The likelihood of large "shoulders" in the oscillogram of syllables beginning with p, c, t is as large as for the same vowels when voiced directly. The most sudden start accompanied the trick used by some singers in building up lung pressure and exploding it through the epiglottis (*coup de glotte*). This investigation has technical importance as it decides the limits to which the background noise cut-off in talking-film recording can operate. Progressive build-up rates which could carry the modulation of an oscillator from zero to 100 per cent in one-twentieth of a second are frequent, though the majority of syllables start more slowly than this.

Marro (1940) has continued his experiment on the intermittent reversal of speech. By means of a valve circuit the wave form of speech is inverted into its image, with reference to the mean line as "mirror", at regular intervals, the intervening portions being left undisturbed. The current so treated is then supplied to a microphone, and it is found that the voice becomes incomprehensible when the inversion takes place 200 to 300 times per second. The object is to produce secrecy and make listening-in to a conversation fruitless.

Burris-Meyer (1940, 41) gives an interesting account of some voice-faking in theatrical productions. Pitch transformations and modulations of the human voice have been used for the Ghost in "Hamlet", Bottom in "A Midsummer Night's Dream" and Ariel in "The Tempest". (Readers may have heard similar tricks in Walt Disney's "Silly Symphonies"). Reverberation may be superimposed on musical records made in a dead studio to give the impression that they are being played in a more reverberant auditorium, like a church.

(ii) *Hearing*

The articulation test—writing down lists of unrelated syllables which the subject hears—has been in use for a number of years for measuring the capacity for interpreting speech sounds. Steinberg and Gardner (1940) show that defective hearing exhibited by audiograms in the form of loudness-loss against frequency within the range of normal speech are sufficiently closely related to the results of articulation tests, but the actual loudness of the received speech is not an adequate criterion of its intelligibility to the deaf person. Dimmick and Olson (1941) have redetermined with a pure-tone oscillator the threshold for discrimination of intensity of a tone. Two tones of different intensity but equal frequency were presented successively. In decibel terms, the mean discrimination fell with increase in the intensity. They point out that it is difficult to compare the results of different investigators in this field, owing to the different technique employed. In this connection Lifshitz (1941) opines that the available psycho-physical data can be summarized by saying that (for

small levels) one can appreciate an increase in loudness only when it is greater than 2 db., and that when the decibel is translated into appropriate units the same holds for the senses of sight and touch.

Using an apparatus based on sound-film technique, Türk and Lottermoser (1940) were able to study the perception of pulses of short duration. Great care had to be taken to eliminate d.c. clicks in turning the sound on and off. It appeared that the ear is very sensitive to starting characteristics of different duration in the sounds which it perceives. Differentiation is of two kinds. With relatively long starting characteristics, the audible difference depends on the envelope, and not much on the quality. When the transient has a short duration only the quality matters. At intermediate time intervals, both growth of amplitude and timbre are perceived.

Watson and Knudsen (1940) have made a painstaking study of the use of selective amplification in hearing-aids. They consider that the "most comfortable level" at which the partially deaf person can hear speech through the apparatus is an important criterion for the functioning of the hearing-aid; the hearing-loss compared to a normal ear (in db.) at each frequency is determined next, and the selectivity of the amplifier adjusted to give the "most comfortable level" over as wide a band as possible. The importance of accommodation by the patient to a system of hearing that he has never experienced before (as with dental plates) is stressed. Until this is secured his judgement as to the improvement may be warped.

(iii) *Musical scales*

Lloyd (1940, 41) gives us a survey of acoustic features of harmony and counterpoint since the sixteenth century. He points out the danger of misreading into a musical composition a scale based on the tuning of a keyboard instrument. The words "just intonation" and "pure scale" are used in two senses. One relates to intonation of concords in contrapuntal writing; the other to juggling with the frequencies of the notes of keyboard instruments, so as to limit the size of keyboards required for compositions which modulate.

Dufton (1940) suggests a compromise "comma" for musical scales. The comma of Pythagoras is the small interval between twelve fifths ($1{\cdot}5^{12}$) and seven octaves (2^7), while that of Didymus is the interval between four fifths ($1{\cdot}5^4$) and two octaves and a third ($1{\cdot}25 \times 2^2$). Neither of these is contained an integral number of times in a single octave, but the mean, Dufton's comma, can run 53 to the octave. The relation of this to just intonation is pointed out.

Allan (1940) has made precise measurements of the frequency of notes produced by several specimens of the Highland bagpipe with a view to determining the musical scale which they follow. The positioning of the side holes seems to be largely haphazard. If any system is followed, it is one governed more by the convenience of the player in putting his fingers over the holes (as it is in all primitive wind instruments), and certainly no diatonic scale is followed. There are, in fact, three-quarter tones among the intervals, but all are approximations, the reason given by the pipers being that this enables scales based on

several key-notes to be employed, or, perhaps, it would be safer to say, that as this makes it fit none at all, it treats every key with equal disdain (or discordance). The coupling between the reed and the pipe is also examined. Allan has succeeded in producing a chanter with more accurately tuned intervals, at the expense of some facility in fingering.

§7. TECHNOLOGY

(i) *Building acoustics*

Maa (1940, 41) continues his theoretical studies of rooms with absorbent materials, calculating the normal modes of cavity resonance and then working out the decrement in terms of the specific acoustic impedance of specimens placed on the walls. Approximate solutions are given for non-uniform distribution of absorbent, with the proviso that diffracted reflections from the walls can be neglected. Experiments were carried out in a model reverberation chamber made of 3-in. thick marble, to which layers of wood or Celotex were fixed, their specific impedances being known. The theory shows that the "absorbing power" of a material is not a quantitative statement of its absorption characteristics. The effect on reverberation depends on the shape of the room and the position of the specimen. Fortunately the term can still be used in an approximate sense. He also discusses the multiple echoes formed between two parallel walls due to a pulse (e.g. footfall) in relation to the normal modes of vibration of the intervening air column.

Mason and Moir (1941) point out the importance of "intimacy" in a cinema, i.e., the proper topographical relation of the sources of sight and sound. Even when the frequency characteristics of the reproducing and the taking auditorium are the same, the impression which an audience gets may not be that of the original. To get true reproduction, there should not be too many reflected sounds coming to an auditor from random directions. Further, the audience will feel the sound quality to be rather above the average when the reverberation time is below that usually quoted as the optimum. As a step towards the ideal *théâtre intime* for talking pictures, they propose an auditorium with a low cut-up ceiling, the height diminishing towards the proscenium side walls, which should be splayed to an angle roughly parallel to the direct sound rays. Strips of acoustic material of coefficient 0·25 to 0·4 are to be placed on the side walls.

Lindahl and H. J. Sabine (1940) and Knowler (1941) have been measuring the transmission of impact noises in buildings. The former worked on a 6-in. concrete slab floor separating two stories of a warehouse. The slab was used both bare and covered with small panels of flooring material. The values for each of these were stated as the transmission level in db. below that of the bare concrete.* Subjectively, the improvement most worth while depended on the reduction of high-frequency components, rather than in overall noise. Knowler was concerned with the transmission of a blow on the floor to different parts of the physics building of the National Physical Laboratory. In a horizontal direction the attenuation was about 0·5 phon per foot run in dull blows, but

only 0·3 phon per foot in sharp blows, which were found to contain a more prominent component of moderate frequency (400 c./sec.) than the low frequency (100 c./sec.) of the "dull" blows. A wall broken by window spaces naturally impedes the propagation of sound more than an unbroken one.

Kawashima (1941 b) describes transmission experiments with non-porous materials in the sound laboratory of Waseda University. Three cases were investigated: (i) a thin flexible partition so held that its stiffness could be neglected; (ii) a thick slab held under elastic restraint. The results of these two verify some formulae deduced by A. H. Davis. The third case (iii), not considered by Davis, is that of an elastically built-in wall of thickness comparable to the wave-length. The corresponding formula for its transmission loss is derived.

Redfearn (1940) discusses some reciprocal relations between source and observer in relation to acoustic fields in which obstacles may be placed. This leads to a study of the practical methods available for reducing the noise-level at a point due to an inexorable source ot another point. This can be done, of course, by barriers or by enclosures, and it then becomes a question whether to enclose the source or the observer, or both. Enclosures round the listener with different-sized apertures in the direct line from source to observer are considered and the variation of noise-level within the enclosure to that at the aperture worked out for various values of the absorption coefficient. This, when compared to free transmission through the aperture to a point at the centre of the enclosure, shows the improvement effected.

(ii) *Microphones*

Clapp and Firestone (1941) have derived a "wattmeter" for measuring sound-energy flow defined by $1/T.\int_0^T pv\,dt$, where p and v are the instantaneous pressure and particle velocity supposed periodic in time T. Alternating electric voltages proportional to p and v measured on pressure and velocity microphones are obtained and applied to the terminals of a suitable wattmeter of the thermocouple type. Before this application the outputs from the microphones have to be amplified and have their phase-shifts corrected. It is possible to switch either microphone alone on to the wattmeter, in such a way as to measure severally the kinetic and potential sound-energy density, thereby adding to the usefulness of the instrument. The pressure unit was formed of two crystal microphones, free from resonance up to 2000 c./sec., mounted in a holder, one at each end of the ribbon (0·0001 in. aluminium, corrugated), which acted as velocity microphone. The resonances of the latter presented a difficult problem. Though its fundamental was low, overtones produced excessive response and phase-shift unless corrected by suitable networks. It was demonstrated that the instrument may be used to measure the absorption coefficient of a specimen terminating a system of stationary waves, if the energy flow and energy density at some point are measured.

E. C. Jordan and Everitt (1941) have been using acoustic models of radio antennae with success. One model is a pipe closed at one end ($\frac{1}{2}$ in. diameter)

and studded with 1/64-in. holes, spaced 1 in. apart. This was set up so as to project above a tiled ground and excited by a tone source, and the field plotted with a uni-directional microphone. With an ordinary microphone, reflections from bodies were very marked. To simulate directional antennae two or more of these were used and the intensity and relative phase of each suitably adjusted. In order to shift the phase without changing the magnitude they used side-tubes in pairs with "trombone slides" in the "transmission line" (i.e., pipe) between the tone source and the radiating end. Data are given comparing the experimental field of the acoustic aerials with theoretical prediction.

(iii) *Reproducers*

Begun (1940) discusses some problems of disc recording. He points out that it is now the practice to use a recording unit which gives constant amplitude of cut below a "transition frequency" (round about 350 c./sec.), and above that to change over to constant-velocity cutting. At a higher frequency still (above 1000 c./sec.) some manufacturers make another change to a compromise recording partly on amplitude and partly on velocity, since the characteristic of a constant-velocity system is to emphasize high-frequency sounds, while it requires such large amplitudes in the bass that it tends to limit the proximity of cutting grooves. Low-frequency background noise is not considered so detrimental owing to the smaller sensitivity of the ear at this end of the spectrum. A pick-up with a low-inertia moving part having a stylus pressure of less than one ounce has been made, to carry a permanent sapphire stylus, which makes it possible to leave out the abrasive which is usually mixed with the shellac to give bearing for the needle on the disc. This reduces considerably the high-frequency background noise. The material used for master recordings is comparatively soft and readily damaged. It must be seen that the frictional force and the stylus pressure remain proportional over a reasonable range, if permanent deformation and excessive wear of the disc do not take place. Cellulose nitrate is the usual material for these, but has a number of disadvantages from the point of view of the man who wishes to cut his own records at home.

Lewis and Hunt (1941) have conducted an elaborate analysis of the distortion in sound reproduction on discs which arises because of the relative shape of recording stylus and reproducing stylus. If these could be identical, the tracing distortion would have disappeared. ("Alas", say they in conclusion, "then our paper would not have been written!") Another distortion in lateral pick-up arises when the narrowing of the groove causes vertical motion in the pick-up. This has been discussed by Fleming (1941), as well as by the former pair of investigators.

Klipsch (1941) describes a low-frequency horn designed to fit into the corner of a room, and to utilize wall and floor reflections to improve the output impedance. It can reproduce sounds of wave-length eight times its mean dimension (about 1 metre) with efficiency and lack of distortion (down to 40 c./sec.). The author uses it in his living-room for listening to broadcast music, in conjunction with a smaller straight horn for frequencies above 400 c./sec. Guests are inclined

to compare the performance with their recollection of original renditions rather than with other reproducing systems.

(iv) *Propeller noise*

Obata and his colleagues (Obata, Matumura, Kanayama and Yosida, 1940; Obata, Yosida and Makita, 1941) have made an extensive study of airscrew noise associated with " flutter " of the blades. Flutter sound consists of three parts, i.e., the sound of rotation (blade frequency passing a fixed point), the vibration of the blades, and combination tones arising between the former. The latter were objective in nature; they could be picked up on an adjustable resonator. When the airscrew was replaced by a thin plate attached to the motor axle at its mid-point it showed that the blade vibrations involved in flutter were torsional modes. Experiments on this model were also conducted under reduced pressure. This reduced both the amplitude of the vibrations and the range of revolutions per minute under which they were excited, showing that aerodynamic forces play a considerable part in the phenomenon. Both in this paper and one by Deming (1940) the polar distribution of noise in relation to the airscrew disc is described.

Problems of the " singing propeller " have been examined by Kerr, Shannon and Arnold (1941) in the light of recent research. They emphasize the importance of the reaction between torsional and flexural vibrations in the mechanism of flutter. The leading edge may be so fashioned as to provide hydrodynamic damping of the flutter (as well as cutting down cavitation at the same time). Sharpening the trailing edge cuts down vorticity, which, if pronounced, encourages vibration.

(v) *Echo sounding*

Mortimer and Worthington (1940) have studied the depth and nature of the bed of Windermere with the Hughes (magneto-striction) technique. Besides an echo from the floor of the lake, others coming from depths up to 13 m. lower were found. By extracting cores of the mud with a special apparatus it has now been confirmed that these reflections come from the clay bands found in the cores and due to seasonal variations in the sediments brought down in post-glacial epochs. Comparison between echo records and core analysis yields values for the velocity of sound in the mud which vary between 1540 and 1680 m./sec These preliminary experiments indicate that echo-sounding equipment may be of considerable use in the service of sublacune geology.

REFERENCES

ALLAN, G. E., 1940. *Phil. Mag.* **29**, 154.
ANDRADE, E. N. DA C., 1941. *Proc. Phys. Soc.* **53**, 129.
BÄR, R., 1940. *Helv. phys. Acta*, **13**, 61.
BARTHOLEMEW, W. T., 1940. *J. Acoust. Soc. Amer.* **11**, 446.
BATE, A. E., 1940. *J. Sci. Instrum.* **17**, 68.
BEGUN, S. J., 1940. *Proc. Inst. Radio Engrs., N.Y.*, **28**, 389.
BENDER, D., 1940. *Ann. Phys., Lpz.*, **38**, 199.
BERANEK, L. L., 1941. *J. Acoust. Soc. Amer.* **12**, 3, 14.

Bhimasenachar, J. and Venkateswarlu, K., 1940. *Proc. Indian Acad. Sci., Bangalore*, **11**, 28.
Bilhuber, P. H. and Johnson, C. A., 1940. *J. Acoust. Soc. Amer.* **11**, 311.
Bolt, R. H. and Brown, R. L., 1940. *J. Acoust. Soc. Amer.* **12**, 31.
Boner, C. P. and Newman, R. B., 1940. *J. Acoust. Soc. Amer.* **12**, 83.
Bopp, E., 1940. *Ann. Phys., Lpz.*, **38**, 495.
Burris-Meyer, H., 1940. *J. Acoust. Soc. Amer.* **12**, 122.
Burris-Meyer, H., 1941. *J. Acoust. Soc. Amer.* **13**, 16.
Clapp, C. W. and Firestone, F. A., 1941. *J. Acoust. Soc. Amer.* **13**, 124.
Colwell, R. C., Stewart, J. K. and Arnott, H. D., 1940. *J. Acoust. Soc. Amer.* **12**, 260.
Cremer, L., 1940. *Akust. Z.* **5**, 57.
Deming, A. F., 1940. *J. Acoust. Soc. Amer.* **12**, 173.
Dimmick, F. L. and Olson, Ruth M., 1941. *J. Acoust. Soc. Amer.* **12**, 157.
Drew, R. O. and Kellogg, E. W., 1940. *J. Acoust. Soc. Amer.* **12**, 95.
Dufton, A. F., 1940. *Phil. Mag.* **32**, 259.
Eagleson, H. V., 1941. *J. Acoust. Soc. Amer.* **12**, 427.
Epstein, P. S., 1941. *Cal. Inst. Tech. Kármán Anniv.* 162.
Eucken, A. and Aybar, J., 1940. *Z. phys. Chem.* B, **46**, 195.
Fairbanks, G., 1940. *J. Acoust. Soc. Amer.* **11**, 457.
Fay, R. D., 1940. *J. Acoust. Soc. Amer.* **12**, 62.
Fleming, L., 1941. *J. Acoust. Soc. Amer.* **12**, 366.
Fox, F. E., 1940. *J. Acoust. Soc. Amer.* **12**, 147.
Fox, F. E. and Rock, G. D., 1941. *J. Acoust. Soc. Amer.* **12**, 505.
Fricke, E. F. and Knudsen, V. O., 1940. *J. Acoust. Soc. Amer.* **12**, 245 and 255.
Ganitta, E., 1940. *Akust. Z.* **5**, 87.
Gregg, E. C., 1941. *Rev. Sci. Instrum.* **12**, 149.
Hartmann, G. K. and Focke, A. B., 1940. *Phys. Rev.* **57**, 221.
Hartmann, J. and Lazarus, P., 1940. *Phil. Mag.* **29**, 140.
Hartmann, J. and Lazarus, P., 1941. *Phil. Mag.* **31**, 35.
Herget, C. M., 1940 a. *Rev. Sci. Instrum.* **11**, 37.
Herget, C. M., 1940 b. *J. Chem. Phys.* **8**, 537.
Herzfeld, K. F., 1941. *J. Acoust. Soc. Amer.* **13**, 33.
Hilton, W. F., 1940. *Phil. Mag.* **30**, 237.
Hunter, J. L., 1941. *J. Acoust. Soc. Amer.* **13**, 36.
Itterbeek, A. van and Mariens, P., 1940. *Physica*, **7**, 125.
Jones, A. T., 1941. *J. Acoust. Soc. Amer.* **12**, 387.
Jordan, E. C., 1941. *J. Acoust. Soc. Amer.* **13**, 8.
Jordan, E. C. and Everitt, W. L., 1941. *Proc. Inst. Radio Engrs., N.Y.*, **29**, 186.
Kanta, Chandra, 1940. *Proc. Nat. Inst. Sci. India*, **6**, 671.
Kanta, Chandra, 1941. *Indian J. Phys.* **15**, 161.
Kastner, L. J., 1941. *Phil. Mag.* **32**, 206.
Kawashima, S., 1941 a. *J. Acoust. Soc. Amer.* **12**, 514.
Kawashima, S., 1941 b. *J. Acoust. Soc. Amer.* **12**, 75, 327 and 332.
Kaye, G. W. C. and Evans, E. J., 1940. *Proc. Phys. Soc.* **52**, 371.
Keller, H. H., 1940. *Phys. Z.* **41**, 386.
Kerr, W., Shannon, J. F. and Arnold, R. N., 1941. *Proc. Inst. Mech. Engrs.* **144**, 54.
Klipsch, P. W., 1941. *J. Acoust. Soc. Amer.* **13**, 137.
Knauss, H. P. and Yeager, W. J., 1941. *J. Acoust. Soc. Amer.* **13**, 160.
Kneser, H. O., 1940. *Ann. Phys., Lpz.*, **39**, 261.
Knowler, A. E., 1941. *Phil. Mag.* **31**, 240.
Krishnan, K. G., 1940. *Proc. Indian Acad. Sci., Bangalore*, **13**, 281.
Labaw, L. W., 1940. *J. Acoust. Soc. Amer.* **12**, 232.
Lenihan, J. M. A. and Richardson, E. G., 1940. *Phil. Mag.* **30**, 357.
Leonard, R. W., 1940. *Rev. Sci. Instrum.* **11**, 389.
Lewis, D. and Tuthill, C., 1940. *J. Acoust. Soc. Amer.* **11**, 451.

LEWIS, W. D. and HUNT, F. V., 1941. *J. Acoust. Soc. Amer.* **12**, 348.
LIFSHITZ, S., 1941. *J. Acoust. Soc. Amer.* **12**, 526.
LINDAHL, R. and SABINE, H. J., 1940. *J. Acoust. Soc. Amer.* **11**, 402.
LINDSAY, R. B., 1941. *J. Acoust. Soc. Amer.* **12**, 378.
LLOYD, L. S., 1940. *J. Acoust. Soc. Amer.* **11**, 440.
LLOYD, L. S., 1941. *J. Acoust. Soc. Amer.* **13**, 170.
MAA, D. Y., 1940. *J. Acoust. Soc. Amer.* **12**, 39.
MAA, D. Y., 1941. *J. Acoust. Soc. Amer.* **13**, 170.
MARRO, M., 1940. *Phil. Mag.* **29**, 205.
MASON, C. J. and MOIR, J., 1941. *J. Inst. Elect. Engrs.* **88**, 175.
MCGINNIS, C. S. and GALLAGHER, C., 1941. *J. Acoust. Soc. Amer.* **12**, 529.
MEYER, E., BUCHMANN, G. and SCHOCK, A., 1940. *Akust. Z.* **5**, 352.
MORSE, B. B., BOLT, R. H. and BROWN, R. L., 1940. *J. Acoust. Soc. Amer.* **12**, 217.
MORTIMER, C. H. and WORTHINGTON, E. B., 1940. *Nature, Lond.*, **145**, 212.
NARASIMHAIYA, R. L. and DORAISWAMI, C. S., 1940. *Indian J. Phys.* **14**, 187.
NORTHRUP, P. A., 1941. *J. Acoust. Soc. Amer.* **12**, 90.
OBATA, J. and KOBAYASHI, R., 1940. *Proc. Phys.-Math. Soc. Japan*, **22**, 691.
OBATA, J., MATUMURA, Y., KANAYAMA, R. and YOSIDA, Y., 1940. *Rep. Aero. Res. Inst. Tokyo*, **15**, pt. 16, 3 and 356.
OBATA, J., YOSIDA, Y. and MAKITA, Y., 1941. *Rep. Aero. Res. Inst. Tokyo*, **16**, pt. 3. 129.
OBERST, H., 1940. *Akust. Z.* **5**, 27.
OKA, S., 1940. *Z. Phys.* **116**, 632.
OSWATITSCH, K., 1941. *Phys. Z.* **42**, 365.
OVERBECK, C. J. and KENDALL, H. C., 1941. *J. Acoust. Soc. Amer.* **13**, 26.
PARKINSON, J. S. and JACK, W. A., 1941. *J. Acoust. Soc. Amer.* **13**, 163.
PETERSEN, O., 1940. *Phys. Z.* **41**, 29.
PIELEMEIER, W. H., SAXTON, H. L. and TELFAIR, D., 1940. *J. Chem. Phys.* **8**, 106.
RAO, B. V. R. and RAMAIYA, D. S. S., 1941. *Phys. Rev.* **60**, 615.
REDFEARN, S. W., 1940. *Phil. Mag.* **3**, 223.
RICHARDSON, E. G., 1940 a. *Proc. Phys. Soc.* **52**, 480.
RICHARDSON, E. G., 1940 b. *Nature, Lond.*, **145**, 841.
RICHTER, C., 1940. *Z. Phys.* **115**, 97.
ROCARD, V., 1940. *Rev. Sci. Paris*, **78**, 209.
ROHLOFF, E., 1940. *Z. Phys.* **117**, 64.
RSCHEVKIN, N. S. and TEROSSIPJANTZ, S. T., 1941. *J. Phys. U.S.S.R.* **4**, 45.
SABINE, P. E., 1941. *J. Acoust. Soc. Amer.* **12**, 317.
SAUNDERS, F. A., 1940. *J. Franklin Inst.* **229**, 1.
SAVIC, P., 1941. *Phil. Mag.* **32**, 245.
SCHAAFS, W., 1941. *Ann. Phys., Lpz.*, **40**, 393.
SCHAEFER, C., 1940. *Z. Phys.* **115**, 109.
SCHÜSTER, K. and MATZ, W., 1940. *Akust. Z.* **5**, 349.
SEIDL, F., 1940. *Z. Phys.* **116**, 359.
SHEAR, S. K. and FOCKE, A. B., 1940. *Phys. Rev.* **57**, 532.
SIBAIYA, L. and NARASIMHAIYA, R. L., 1941. *J. Mysore Univ.* **1**, 133.
SIEG, H., 1940. *Elekt. Nachr.-Tech.* **17**, 193.
STEINBERG, J. C. and GARDNER, M. B., 1940. *J. Acoust. Soc. Amer.* **11**, 270.
SURYAPRAKASAM, V., 1940. *Proc. Indian Acad. Sci., Bangalore*, **12**, 341.
THOMPSON, L., 1940. *J. Acoust. Soc. Amer.* **12**, 198.
TÜRK, W. and LOTTERMOSER, W., 1940. *Akust. Z.* **5**, 324.
TYTE, L. C., 1940. *Phil. Mag.* **30**, 173.
UNWIN, J. J., 1941. *Proc. Roy. Soc.* A, **178**, 153.
VERMEULEN, R., 1940. *Philips Tech. Rev.* **5**, 36.
WALLER, MARY D., 1940. *Proc. Phys. Soc.* **52**, 452.
WATSON, N. A. and KNUDSEN, V. O., 1940. *J. Acoust. Soc. Amer.* **11**, 406.
WILLARD, G. W., 1941. *J. Acoust. Soc. Amer.* **12**, 438.
ZAHRADNICEK, J. and NESPOR, R., 1940. *Phys. Z.* **41**, 419.

EXPERIMENTAL STUDY OF X-RAY SCATTERING IN RELATION TO CRYSTAL DYNAMICS

By KATHLEEN LONSDALE,
Royal Institution, London

§ 1. INTRODUCTION

EXPERIMENT has shown that when a beam of X rays falls upon a crystal, even though it may be limited to a small range of angles of incidence, secondary X radiation emerges in all directions from the crystal. In general, the total intensity of secondary radiation is small compared with the intensity of the incident beam, the greater part of which proceeds through the crystal and emerges from it to form the *central* or undeviated beam. The secondary radiation, although emitted in all directions, is not uniformly distributed, it is not all of the same wave-length, nor is it uniformly polarized. It is, in fact, produced or influenced by a number of processes which are taking place simultaneously, or practically simultaneously, either at the surface of, or within, the irradiated crystal. In order to investigate any one of these processes, it is necessary to understand how to select for examination just that part of the secondary radiation which will give the required information. The rest must be eliminated; but before it can be properly eliminated it must be recognized and possibly studied.

During the past thirty or forty years many different kinds of investigation have shown that the secondary radiation emerging from an irradiated crystal may consist of: (i) fluorescent radiation, which is less penetrating and of greater wave-length than the primary radiation and is unpolarized unless, before emerging, it has suffered internal reflection; (ii) Compton scattering, which forms a background of diffuse, incoherent radiation of wave-lengths slightly longer than those of the primary radiation; (iii) Laue or Bragg reflections, which are intense, sharply defined and strongly polarized beams of scattered radiation of the same wave-length as the incident beam; (iv) regions of relatively strong, but diffuse, scattered radiation, the wave-length of which is not measurably different from that of the incident beam (Faxén-Waller scattering); (v) a general scattering of weak diffuse radiation of the same type as (iv) (Debye-Jauncey scattering). Practically the whole of the scattered radiation is polarized.

These different kinds of radiation vary in relative importance according to the wave-length (or wave-lengths), divergence and polarization of the incident beam, the nature, temperature, orientation and condition of the crystal, and the angle of scattering. Their effects will also vary with the method of detection used, and the results may be complicated by multiple scattering, tertiary

radiation or indirect excitation ("unweganregung": Renninger, 1937; Blanc and Weigle, 1937; v. Laue, 1941). In the following paragraphs the properties of each of the above five types of scattering will be briefly described, references being given only where these properties have been the subject of recent study.

In addition to all the above, however, it must be remembered that air scattering, scattering from impurities and from various parts of the apparatus, cosmic rays and other extraneous effects, if not adequately eliminated, can obscure or modify the observations (Laval, 1939 b; McNatt, 1939), while the presence of discontinuities in absorption by the photographic emulsion or ionization chamber gas must also be allowed for (Baxter and Brentano, 1937; Brentano, Honeyburne and Berry, 1939).

§ 2. FLUORESCENT RADIATION

Fluorescent radiation originates in the ionization, by x rays of sufficient energy, of the atoms in the crystal, and in the liberation of energy as those atoms regain their normal states. Its wave-lengths are those characteristic of the atoms of the scatterer (K, L, M....), and they are, of course, independent of those of the primary radiation, with the essential proviso that unless the primary radiation contains short enough wave-lengths, the characteristic fluorescence rays will not be excited. In other words, the exciting primary beam will only produce K-, L-, M-.... fluorescence rays if it contains radiation of frequency equal to or greater than those frequencies characteristic of the K-, L-, M-.... absorption edges of the elements in the crystal. The fluorescent radiation is unpolarized and is uniformly distributed in all directions. The fluorescence yield, or ratio of the number of atoms which emit Q radiation to the number from which Q electrons have been ejected ($Q = K, L, M \ldots$.) is small for elements of low atomic number and only approaches unity for heavy elements. *In experiments on X-ray scattering from crystals, fluorescent radiation is therefore only of importance when the crystal contains heavy elements and the primary beam contains sufficiently short wave-lengths.* When it does occur, however, it can be much stronger than the general scattered radiation (Jauncey, 1939), even though the responsible primary wave-length is present in relatively small quantity, as for example in the form of a harmonic of a monochromatic beam (McNatt, 1939); it may also emerge from a single crystal with greater or less intensity in certain directions, owing to internal reflection (Kossel and Voges, 1935; Kossel, 1937; v. Laue, 1935, 37; Borrmann, 1936; Voges, 1936).

Long fluorescent radiation (M, N....) will usually be absorbed either by the surface layers of the crystal or by the surrounding air. K- or L-fluorescent radiation can be eliminated: (i) by reducing the peak voltage so that the primary beam does not contain any wave-lengths short enough to excite the unwanted radiation; this will be done, of course, at the expense of primary intensity; (ii) by inserting in the path of the *secondary* radiation under examination a thickness of aluminium, or other absorber, sufficient to remove the less penetrating radiation. This second method is only satisfactory when the main part of the primary beam and of the scattered radiation consists of penetrating radiation

of short wave-length; (iii) if the primary beam is monochromatized and consists mainly of long wave-length radiation, but contains short-wave harmonics, a differential method of elimination may be used, the absorbing aluminium being placed in the *primary* beam (McNatt, 1939). The fluorescence coefficient is, apart from the Kossel effect, independent of crystal orientation (Baltzer and Jauncey, 1939), and also of the state of motion of the atoms in the crystal. Fluorescent radiation may be particularly troublesome if a heavy impurity is present in a light scatterer.

§3. COMPTON SCATTERING

Spectroscopic examination or analysis by absorption of the secondary x radiation from light elements, irradiated by a monochromatic primary beam, shows that, in addition to scattered radiation of the same wave-length as that of the incident beam, there is present radiation of slightly longer wave-length, which is not, however, fluorescent radiation. This *modified* radiation was correctly interpreted by Compton in terms of the old quantum theory and has since received also a satisfactory wave-mechanical explanation (Compton and Allison, 1935).

In the original form of the theory the scattering electrons, supposed to be free and initially at rest, take up energy from the incident x-ray photons and recoil, the photons being scattered with reduced energy which results in an increase of their wave-length from λ to λ', where

$$\lambda' - \lambda = \frac{2h}{mc} \sin^2 \frac{\chi}{2} \qquad \ldots\ldots(1)$$

(h = quantum of action, m = mass of electron, c = velocity of light, χ = angle of scattering relative to primary radiation). The electrons in an atom are not, however, free or at rest, and part of the imparted energy of the photon must be used in removing the electron from its initial orbit to the continuum outside the atom, after which it is free to recoil with whatever energy remains. The observed Compton line, consequently, is not sharp but broad and diffuse, with a definite intensity structure; its maximum intensity occurs at a wave-length λ' given by

$$\lambda' - \lambda = \frac{2h}{mc} \sin^2 \frac{\chi}{2} - D\lambda^2, \qquad \ldots\ldots(2)$$

where D is a constant which depends upon the numbers and binding energies of the electrons in the scatterer. This *shift defect*, $D\lambda^2$, was first found experimentally by Ross and P. Kirkpatrick (1934) and later verified by DuMond and H. A. Kirkpatrick (1937). It was explained as a quantum effect by the former experimenters, and more completely on wave-mechanical theory by Bloch (1934).

The *breadth* of the Compton line has been shown, both experimentally (H. A. Kirkpatrick and DuMond, 1938, give previous references) and theoretically (Hicks, 1937, gives previous references), to vary as a function of the scattering angle, the primary wave-length, the scattering material and its physical state

(gaseous, liquid or solid). Where the scattering electron was originally bound or in motion, the normal Compton wave-length shift will be increased or reduced by an amount which may be regarded as a Doppler effect due to the initial momentum or proper motion of the electron; and thus the breadth and intensity distribution of the Compton line due to an assemblage of electrons, such as those belonging to the atoms or molecules of a gaseous or solid scatterer, will, if properly interpreted, give information concerning the initial velocity distribution of the electrons. A good historical and critical review of the early work in this field has been given by DuMond and Kirkpatrick (1937), who have also tried to determine whether there is any anisotropy of the electron momenta in a scatterer composed of orientated graphite flakes (DuMond, 1932; DuMond, Kirkpatrick and Alden, 1932; Kirkpatrick and DuMond, 1938). So far they have found no dependence of the shape of the Compton line upon the crystal orientation. It should be emphasized perhaps that these experiments are extremely difficult, tedious adjustment of apparatus and photographic exposures of up to more than 2000 hours being necessary; so that, although in theory much information of great value and interest might be expected from a careful study of Compton-line profiles (Hicks, 1940; Duncanson and Coulson. 1941; Coulson and Duncanson, 1942), in practice the x-ray method is still in its infancy. DuMond and Kirkpatrick (1937) believe, however, that they have definite evidence from their experiments that the electron velocities in solid carbon are some 60 % higher than those in the free atom. According to theory, the Compton line should be limited on the short-wave length side by a discontinuity corresponding to a loss of energy of the primary photon equal to the binding energy of the scattering electrons (Sommerfeld, 1937). Such a discontinuity has been observed by Kirkpatrick and DuMond (1938), but its interpretation on the above lines is uncertain. Between the modified and unmodified lines in the spectrum of scattered radiation there might also be expected a series of weak, sharp lines due to scattering by electrons which have only absorbed sufficient energy to transfer them to a state of higher energy within the atom. This kind of Raman scattering in the x-ray region has not yet been observed.

Any information concerning the conduction electrons in metals is particularly welcome, because the ordinary x-ray methods are not likely to be of much value in this respect; these electrons contribute strongly to the Compton scattering (Zener, 1935; Debye, 1937; Scharwächter, 1937), and experiments on beryllium (DuMond, 1929, 33) have proved that their velocities, as deduced from the Compton-line profiles, are in good agreement with calculations based on the Fermi statistics. These electron velocities are practically independent of temperature, and a change of temperature of the scatterer should not, therefore, affect the Compton scattering.

Ingenious cloud-track experiments have shown that the time interval between the incidence of the primary photons and the emission of scattered photons and recoil electrons is less than 10^{-4} second, and have also verified the angular relations given by the photon theory to within $\pm 20°$ (Shankland, 1937, gives many previous references).

The *intensity* of Compton scattering varies markedly with scattering angle, wave-length of primary radiation and the nature of the scatterer; in terms of these parameters the accepted expression for the modified intensity is

$$I_{\text{mod.}} = I_e R[Z - \Sigma(f_{kk}^2) - \Sigma\Sigma_{k\neq l}(f_{kl}^2)](1+K), \qquad \ldots\ldots(3)$$

where $I_e = I_0 \frac{e^4}{2r^2m^2c^4}(1+\cos^2\chi)$ is the intensity of (Thomson) scattering from a free electron, measured at distance r and scattering angle χ, I_0 being the primary intensity. $I_{\text{mod.}}$ is the intensity of Compton scattering, under the same condition, from an atom having Z electrons; f_{kk} being the electronic structure factor due to any one of them. [The essential incoherence of the Compton scattering is shown by the fact that the summation $\Sigma(f_{kk}^2)$ is taken over intensities and not amplitudes of scattering from the contributing electrons.] R is the Breit-Dirac factor $(\lambda/\lambda')^3$, introduced to satisfy wave-mechanical considerations, and $\Sigma\Sigma_{k\neq l}(f_{kl}^2)$ is the Waller-Hartree correcting term, which expresses the operation of the Pauli exclusion principle, the summation being taken over all pairs of electrons with the same spin, and $k \neq l$. $K = \left(\frac{h}{mc}\right)^2 \frac{(1-\cos\chi)^2}{\lambda\lambda'(1+\cos^2\chi)}$ is the Klein-Nishina correction, which is of theoretical interest in that it represents the only part of the Compton scattering that is not polarized.

Since in any experiments involving the study of intensities of diffuse (temperature) scattering it is essential that the Compton scattering should be allowed for, it will be of interest to estimate the values of the various terms under changing conditions, and to compare the final result with the average unmodified intensity $(\Sigma f_{kk})^2$. Tables of $\Sigma(f_{kk}^2)$ as well as of Σf_{kk} are given as an appendix by Compton and Allison (1935). Waller and Hartree (1929), and Harvey, Williams and Jauncey (1934) have given a few values of $\Sigma\Sigma_{k\neq l}(f_{kl}^2)$. The Klein-Nishina factor is negligible in x-ray regions and need only be considered for γ-rays, while R, though not negligible, is readily calculated with sufficient accuracy from equation (1).

Let us take the case of sylvine (KCl) with Mo $K\alpha$ ($\lambda = 0{\cdot}711$ kx.*) as the primary radiation. The critical K-absorption wave-lengths of K and Cl are 3·431 and 4·384 kx., so that fluorescent radiation of wave-lengths 3·447, 3·736, 4·394, 4·720 kx. will be excited, but such long waves will certainly be absorbed before they reach the detecting apparatus. These absorption wave-lengths are also sufficiently far removed from the primary wave-length for dispersion corrections to be unnecessary (such a correction would, however, be advisable if Cu $K\alpha$ radiation, $\lambda = 1{\cdot}539$ kx., were used; Hönl, 1933; Brindley and Ridley, 1938). The intensity of Compton scattering from the molecule will be given with reasonable accuracy by

$$I_{\text{mod.}} = I_e R[36 - \Sigma(f_{kk}^2)_{K^+} - \Sigma(f_{kk}^2)_{Cl^-} - 2\Sigma\Sigma_{k\neq l}(f_{kl}^2)_A],$$

the Waller-Hartree correction term for argon being a good approximation to

* The kx unit, in terms of which the 200 spacing of NaCl is 2·81400 and the wave-length of Cu$K\alpha$ is 1·537395, is about 0·2% greater than the Ångström unit A. See Lipson and Riley (1943) and Wilson (1943).

that for either K^+ or Cl^-. The intensity of *average* total unmodified scattering per molecule (all effects due to the crystal lattice and temperature movements being smoothed out) is

$$\bar{I}_{\text{unmod. total}} = I_e[(\Sigma f_{kk})^2_{K^+} + (\Sigma f_{kk})^2_{Cl^-}],$$

while the *average* unmodified background scattering per molecule is

$$\bar{I}_{\text{unmod. background}} = I_e\{[(\Sigma f_{kk})^2(1-e^{-2M})]_{K^+} + [(\Sigma f_{kk})^2(1-e^{-2M})]_{Cl^-}\}.$$

M, which is sensibly the same for K^+ and Cl^-, has been given by James and Brindley (1928) as

$$M = \left(\frac{\sin\chi/2}{\lambda}\right)^2 . 8\pi^2\left\{7{\cdot}25\times10^{-5}.T + \frac{0{\cdot}1075}{T} - \frac{56}{T^2}\right\}.$$

T being the absolute temperature.

Table 1. Compton and unmodified scattering from KCl for Mo $K\alpha$ primary radiation

χ	$\frac{\sin\chi/2}{\lambda}$	R	$\Sigma(f_{kk}^2)$		$\Sigma\Sigma(f_{kl}^2)$	$\frac{I_{\text{mod.}}}{I_e}$	$(\Sigma f_{kk})^2$		$\frac{\bar{I}_{\text{u. total}}}{I_e}$	$\frac{\bar{I}_{\text{unmod. background}}}{I_e}$			
			K	Cl	A		K	Cl		86°	293°	550°	665° K.
0	0	1·0	18	18	0	0	324	324	648	0	0	0	0
10	0·123	1·0	14·0	12·3	0·3$_5$	9·0	253	207	460	8·0	23·0	41·9	50·1
20	0·245	0·994	9·8	8·8	0·7	15·9	148	108	256	17·3	47·4	80·6	93·7
30	0·364	0·986	7·7	6·9	1·0	19·1	90·3	70·6	161	23·2	58·6	91·7	102·6
40	0·482	0·977	6·4	5·6	1·2	21·1	62·4	54·8	117	27·9	64·1	90·2	97·2
50	0·595	0·965	5·2	4·4	1·4	23·0	50·4	42·3	93	31·4	65·0	82·9	86·8
60	0·704	0·950	4·4	3·6	1·4	23·9	41·0	32·5	73·5	32·3	60·0	70·4	71·8
70	0·806	0·935	3·7	2·9	1·3	25·1	33·6	25·0	58·6	31·2	52·3	57·6	58·2
80	0·905	0·920	3·1	2·5	1·1	25·9	28·1	19·4	47·5	29·3	44·6	47·2	47·4
90	0·996	0·904	2·7	2·2	1·0	26·3	23·0	15·2	38·2	26·2	36·9	38·1	38·2
100	1·079	0·889	2·4	2·0	0·8	26·5	18·5	11·6	30·1	22·4	29·5	30·1	30·1

It will be seen from table 1 that the Compton and unmodified background scatterings are comparable at liquid-air temperatures, the Compton scattering becoming of increasing importance as the scattering angle increases beyond 90°. The unmodified background scattering becomes more prominent as the temperature rises. Similar calculations (table 2) show that the Compton scattering is relatively more important for light elements.

Table 2. Ratio of Compton to unmodified scattering for various elements

$\frac{\sin\chi\cdot2}{\lambda}$	$\frac{I_{\text{mod.}}}{\bar{I}_{\text{unmod. total}}}$ (Waller-Hartree and Breit-Dirac factors neglected)			
	Li	Ne	A	Xe
0	0	0	0	0
0·5	2·5	0·7	0·2	0·04
1·0	33·3	3·7	0·9	0·15

These facts have been qualitatively verified by many experimenters. At large angles and high temperatures the scattering from a crystal approaches that to be expected from the corresponding gas.

§ 4. SELECTIVE REFLECTION

Under certain conditions the scattered radiation can form relatively intense beams in definite directions; these are the Laue or Bragg reflections, and the required conditions for any reflection are those that will satisfy the well-known Bragg relation $n\lambda = 2d \sin\theta$. If the incident beam is not too divergent, the crystal not too small (say $>10^{-5}$ cm. linear dimensions) and the atomic arrangement regular, these reflections are very sharply defined. Under favourable conditions the angular divergence of the reflected beam—half-width at half-maximum intensity—can be as little as 3 seconds of arc (Kirkpatrick and Ross, 1933), but it may be as much as several degrees if the crystal is deformed or composed of an irregular mosaic. For crystals containing only light atoms these reflected beams, which are due to constructive reinforcement of the coherent scattered beams from the regularly arranged groups of electrons, are found mainly around the forward direction of the central beam and are of negligible intensity in the backward direction, where the scattering is all, or nearly all, incoherent. For crystals containing heavy elements the reflected beams may be intense in the backward directions also. This falling-off of intensity with scattering angle, especially noticeable for atoms of low atomic number, is due partly to destructive interference of the scattered beams from electrons within each atomic group, and this factor has been not only calculated, but measured experimentally for many atoms and ions, and is known as the atom form factor, the atomic scattering factor or the atomic structure factor. It is necessary to distinguish between the atomic scattering factor referred to the nucleus ($f = \Sigma f_{kk}$) and the atomic scattering factor referred to the mean position of the nucleus in the crystal lattice ($f_T = fe^{-M}$). The Debye-Waller theory, based on the Born-Kármán theory of lattice dynamics, has shown that M is a function of scattering angle, primary wave-length, temperature, crystal nature and orientation. Experiment has shown that even at absolute zero $f_{T=0} < f$, and that the atom does, therefore, possess zero-point energy; but it should be understood that the possession of zero-point energy does not imply an irregular motion of the atoms at absolute zero, but only the wave-mechanical probability of their being near, but not at, their mean lattice positions.

As in the case of all true scattering, the reflected beams are polarized, in and at right angles to the plane of the incident and reflected beams, and these two components decrease in intensity at a different rate as the scattering angle increases. This polarization is of particular importance when a strong Bragg reflection is to be used as a monochromatic primary beam in order that the scattering from a second crystal may be studied.

For the Bragg relation $n\lambda = 2d \sin\theta$ to be satisfied, either λ or θ must in general be variable for any given crystal. The incident beam must therefore be heterogeneous, or the crystal orientation must be adjustable. In the first case

(λ variable) the reflections are usually known as *Laue* reflections, and occur at grazing angles of incidence θ_L; in the second case (λ fixed) the reflections are called *Bragg* reflections, and occur at grazing angles of incidence θ_B.

The whole question of the selective reflection of x rays from crystals, methods of observation, correction for absorption and extinction (dependence upon crystal perfection and size), and interpretation in terms of the distribution of electron density in the atoms and in the crystal lattice, has been adequately dealt with in many articles and books (some of which are given under General References). In the following discussion we shall be mainly concerned with modifications of the reflections which have recently been experimentally observed as a result of temperature variation or which are due to an anisotropy of the electron distribution about the atomic nuclei or about the mean atomic positions in the crystal lattice.

Early quantitative experiments established the validity of the Debye-Waller formula, $I_T = Ie^{-2M}$, in the case of cubic crystals at temperatures not too near to the melting-point. Here

$$M = \frac{6h^2T}{mk\Theta^2} \cdot \frac{\sin^2\theta}{\lambda^2} \cdot Q\left(\frac{\Theta}{T}\right), \qquad \ldots\ldots(4)$$

where m is the mass of the atom, k the Boltzmann constant, Θ the Debye temperature and $Q\left(\frac{\Theta}{T}\right)$ a quantization factor which has been tabulated and which does not deviate appreciably from unity unless $\frac{\Theta}{T} > 1$ (see Blake, 1933, for a good critical account of early work, and the accompanying article by Professor Max Born for an outline of the derivation of the formula). In considering the temperature variation of the Laue or Bragg intensities from a crystal containing more than one kind of atom, allowance must be made for the fact that each atom has its own temperature factor M_K. Zener and Bilinsky (1936) have also pointed out that Θ is not independent of T, and have obtained improved agreement at high temperatures by making a correction for this fact. M is sometimes expressed in the form

$$M = 8\pi^2 \frac{\sin^2\theta}{\lambda^2} \overline{u}_x^2, \qquad \ldots\ldots(5)$$

where $\overline{u}_x^2$ is the mean square displacement of the atoms, in any arbitrary direction x, from their mean positions; $\overline{u}_x^2$ is also expressible in terms of the absolute temperature T, the mass of the atom and the elastic constants of the crystal; formula (5) then becomes

$$M = 8\pi^2 \frac{\sin^2\theta}{\lambda^2}\left(\beta T + \frac{\gamma}{T} + \frac{\varkappa}{T^3}\right), \qquad \ldots\ldots(6)$$

which was the form given in § 3 of this *Report*, as used by James and Brindley. From formula (5) values of $\overline{u}_x^2$ have been experimentally determined for some cubic crystals, and these are in excellent agreement with calculations based on recent determinations of the positions of the maxima in the frequency spectra

of those crystals (Kellermann, 1940, 41; Iona (Jr.), 1941; Lonsdale, 1942c). Given $\overline{u_x^2}$ for $T>\Theta$, Θ can be calculated from the formula $\overline{u_x^2}=\frac{3h^2TQ}{4\pi^2mk\Theta^2}$, where $Q\sim1$.

In all the earlier experiments it was assumed that the mean-square displacement in any arbitrary direction was one-third of the mean-square total displacement, but this assumption was purely one of convenience, and it will be shown later that even for cubic crystals the atomic displacements in certain directions in the lattice are much greater than in others. That the lattice vibrations in certain hexagonal crystals are not isotropic was first suggested by observations on Zn made by Brindley and Spiers (1935). Zener (1936; cf. also Kohler, 1939) then pointed out that in anisotropic crystals the temperature factor will be dependent upon the orientation of the reflection plane, and his formula

$$\overline{u_\psi^2}=u_0\cos^2\psi+\overline{u_{90}^2}\sin^2\psi$$

or

$$M_\psi=M_0\cos^2\psi+M_{90}\sin^2\psi,$$

where ψ is the angle between the c crystallographic axis and the normal to the reflection plane, has been used by Brindley (1936: Zn and Cd), Wollan and Harvey (1937: Zn), Wollan (1938; Cd–Ag alloy), Howard (1938: Zn–Cu alloy), and Brindley and Ridley (1938: Mg; 1939: Cd) to determine the anisotropy of the lattice vibrations in the various hexagonal crystals named. This anisotropy in turn implies a dependence of the characteristic Debye temperature Θ upon ψ (cf. Grüneisen and Goens, 1924; Born and Göppert-Mayer, 1933, p. 689). Since the value of M increases with $\sin\theta/\lambda$, it follows that the effect of lattice vibrations in decreasing the intensity of selective reflection will be more obvious at large than at small angles, and also for short than for long wave-lengths, a fact that accounts for the very remarkable changes that sometimes take place in Laue photographs as T varies. Brindley and Spiers' measurements, however, showed an intensity discrepancy even for *small* angle reflections in the case of Zn, and this was confirmed by Miller and Foster (1936), Miller (1937) and Wollan and Harvey (1937). Correction for the differential lattice vibrations only rendered this anomalous result more obvious, and measurements of diffuse scattering from Zn, which will be referred to again in a later section, also confirmed the conclusion that the electron atmosphere in Zn crystals is distorted so that the atom, even at rest, is somewhat spheroidal in shape, with its long axis along the c crystallographic axis. This distortion is still further accentuated by the anisotropic lattice vibrations. The dependence of f upon ψ which is shown by these experiments persists to $\sin\theta/\lambda=0{\cdot}6$, indicating that several electrons in Zn are associated with the asymmetry, which may therefore extend to other than the valency electrons. In the case of the alloys examined (Cd–Ag, 20% Ag, ϵ phase; Zn–Cu, 15% Cu, ϵ phase), the apparent difference between $f_{\psi=0^\circ}$ (parallel to c axis) and $f_{\psi=90^\circ}$ (normal to c axis) approaches very large values, again indicating an asymmetry of the electron atmosphere about the nucleus; but in these cases the distortion is such as to make the atom an *oblate* spheroid,

with the short axis along the c crystallographic axis; and the lattice vibrations are also anisotropic with the root mean-square displacements normal to c greater than those parallel to c. It is naturally interesting to relate these anisotropies to the departure of the axial ratio from the ideal value 1·633:

Table 3

Crystal	c/a	$(\overline{u_a^2})^{\frac{1}{2}}$ $(\overline{u_{90}^2})^{\frac{1}{2}}$ (290° K.)	Comparison of atom form factors, at rest
Zn	1·856	0·153 0·091	$f(\psi=0°, \sin\theta/\lambda=0·203)$ 20·4; $f(\psi=90°, \sin\theta/\lambda=0·217)$ 23·6
Cd	1·886	0·182 0·118	$f(\psi=0°, \sin\theta/\lambda=0·179) < f(\psi=90°, \sin\theta/\lambda=0·194)$
Mg	1·624	0·125 0·130	
Cd–Ag	1·56	$\overline{u_a^2} < \overline{u_{90}^2}$	$f(\psi=0°, \sin\theta/\lambda=0·21) > f(\psi=90°, \sin\theta/\lambda=0·19)$
Zn–Cu	1·56	0·138 0·153	$f(\psi=0°, \sin\theta/\lambda=0·234)$ 25·2; $f(\psi=90°, \sin\theta/\lambda=0·207)$ 16·5

That the value of the axial ratio is not conditioned by the lattice vibrations is proved by the fact that when the crystals are cooled to liquid air temperatures the atomic vibrations are only about half as large as at room temperatures, but the axial ratio changes very little (Howard, 1938). The asymmetry of the lattice vibrations appears to be an effect rather than a cause of the lattice asymmetry. Nor is the asymmetry of the electron distribution about the atomic nuclei in these crystals an isolated example of such anisotropy. Brindley and Ridley (1937) have shown that the intensities of selective reflection from Rb indicate a concentration of electron density at the intersections of (0002) and (11$\bar{2}$0) planes, half-way between (10$\bar{1}n$) planes.* Lark-Horovitz and Ehrhardt (1939, 40), James and Johnson (1939; see for previous references), Johnson (1940) and others have reported electron-diffraction and x-ray intensity data for ZnO, a tetrahedral hemimorphic crystal, which together show that after the application of a temperature correction there remains an asymmetry of the electron atmosphere about each atom corresponding to a measure of homopolar linkage between nearest Zn and O neighbours along the c axis. There is also a

* The conventions adopted throughout this paper in respect of indices and brackets are as follows:

$(h_1h_2h_3)$ are the Miller indices of a crystal plane in all systems except the hexagonal.

$(h_1h_2\overline{h_1+h_2}h_3)$ are the indices of a plane in the hexagonal system.

$h_1h_2h_3$ is the reflection from the $(h_1h_2h_3)$ plane.

(200), say, represents a set of planes parallel to the cube faces of a cubic crystal, but having half the normal spacing.

200 is the reflection from those planes; (that is, it is the second-order reflection from the (100) planes).

$\{h_1h_2h_3\}$ is a form, or set, of planes or reflections; (thus $\{111\}$ would mean (111) (1$\bar{1}$1) (11$\bar{1}$) ($\bar{1}$11) (1$\bar{1}\bar{1}$) ($\bar{1}$1$\bar{1}$) ($\bar{1}\bar{1}$1) and ($\bar{1}\bar{1}\bar{1}$), in the cubic system; $\{400\}$ would mean (400) (040) (004) ($\bar{4}$00) (0$\bar{4}$0) and (00$\bar{4}$), or the set of reflections from those planes).

$[h_1h_2h_3]$ is the direction of the diagonal, from the origin to the opposite corner of a parallelopiped, the edges of which are h_1a_1, h_2a_2, h_3a_3; a_1, a_2, a_3 being the edges of the unit cell.

[011] will, in the cubic system, be the direction of the normal to the (011) plane, and similarly for all other directions.

suggestion that in metallic Li (body-centred cubic) there may be a maximum of electron density half-way along the cube axes, more pronounced at liquid air temperatures (Griffith and Harris, 1941); moreover, the atom form factor found for metallic Li, after correction for temperature vibrations, is much smaller than that of Li^+ in LiF, indicating a not unexpected looseness of the conduction electrons. Finally, Brill, Hermann and Peters (1942) have proved that in metallic Mg about 1·75 electrons per Mg atom are uniformly spread over the whole lattice, and there are also in the lattice definite points of electron concentration and of electron deficit. The importance of considering such effects as a corollary to the study of lattice vibrations will be emphasized later, when it will be shown that a special experimental investigation has been necessary to decide whether the 222 "forbidden" reflection of diamond is the result of a particular kind of lattice vibration or of a polarization of the electron cloud about each carbon nucleus.

The asymmetry of the actual atomic movements is likely to be most pronounced for "layer" structures, and Hughes (1941) has found it necessary, in investigating the structure of melamine, a monoclinic organic compound, to make a differential allowance for the effect of the anisotropic temperature vibrations in reducing the intensity of reflection from planes at various angles to the mean plane of the molecule. On the other hand, the "rotation" of groups in the crystal will sometimes give high temperature polymorphs which are more symmetrical than the normal arrangement of the groups when not subject to rotation (Hendricks, Posnjak and Kracek, 1932; Finbak and Hassel, 1935, 37; Finbak, 1939)

At the best, however, measurements on the temperature changes of intensity of selective reflection only give average root mean-square amplitudes in various directions. In the next section it will be seen that X-ray measurements of the diffuse scattering promise to give information concerning the whole frequency spectrum.

One factor which strongly affects the intensity of selective reflection is the *extinction*, which is dependent upon the size and perfection of the crystal blocks. If, therefore, the heat or other treatment to which a crystal is subjected varies the state of perfection or mosaic character of the crystal, the Laue or Bragg reflections will change, not because of any real variation of the scattering power of the atom in the lattice, but because of a variation in extinction. This fact, well known to early experimenters, has been underlined by the recent work of Goetz and Jacobs (1935, 37), who have found that whereas the formula $I_T = Ie^{-2M}$ holds for Bi crystals from about 25°–450° K., there is at about 470° K. a minimum in the I_T curve, after which the intensity *increases* up to the melting-point, 545° K. (highest temperature of observations, 538° K.). Moreover, there is a hysteresis of I_T/I over the whole temperature region, which is dependent upon the thermal history of the crystal and which may be attributed to a temperature dependence of the secondary extinction. Dawton (1937) has found a similar hysteresis in reflecting power/temperature curves for sodium, especially for strong reflections.

Many observations from 1931 onwards show that when a quartz crystal, cut in various directions, is subjected to piezo-electric oscillations of different modes, an increase of selective reflection takes place. The increase of Laue spot intensity is noticeably confined to reflection from the interior of the crystal and to those parts of the crystal where maximum displacement takes place (Nishikawa, Sakisaka and Sumato, 1933, 34; Colby and Harris, 1934; Klauer, 1935). The surface layers of the crystal are already sufficiently imperfect to be reflecting without extinction; this sometimes results in a doubling of the Laue spots. The markedly inhomogeneous deformation which occurs at the loops of the piezo-electric oscillation reduces the extinction at definite strata within the crystal and causes an increase of intensity which may show itself as a transformation from the double spots given by a non-oscillating crystal to uniformly intense single spots or to multiple spots, according to the mode of vibration employed. Similar, though less definite, observations have been made on other crystals (Fox and Fraser, 1935; Bertsch, 1936), and it has even been found that a strong *direct* electric field will produce such inhomogeneous strains that a considerable diminution in extinction results (Němejcová and Brož, 1938; Kakiuchi, 1938). Early experiments suggested a change in the intensity and spread of the central beam after its passage through piezo-electrically oscillating quartz, but this has been negatived by several more recent investigations, which also indicate that there is no change in the average diffuse scattering (Jauncey and Deming, 1935*a*; Jauncey and Jaques, 1936; Fox and Frederick, 1938; Jauncey and Bruce, 1938; Weigle and Smith, 1942). Temperature effects, if any, are too small for observation (Fox and Stebbins, 1939).

§5. FAXÉN-WALLER SCATTERING

(i) *Introduction*

The certain existence of regions of relatively strong diffuse scattering in the neighbourhood of positions of actual or potential selective reflection (that is, when the Bragg relation is nearly but not exactly satisfied) has only been established experimentally within the last five years. The principal credit in this respect is due to J. Laval (1938, 39), whose classical researches using monochromatic radiation were not only the first but easily the most extensive in this field.

Using *monochromatic radiation* and a rotating or oscillating crystal the regions of strong scattering take the form of a diffuse background to the Bragg spots, the shape, intensity and extent of which vary with the temperature and with the spot under consideration. For a powdered crystal each line is diffusely broadened, again to a variable extent; while in the case of a stationary crystal a pattern of diffuse spots, streaks and areas may appear even when there is no selective reflection at all (plates 1 *a*, 2 *a*, 3 *a*). The spots invariably appear, however, in a position such that the Bragg relation is approximately satisfied for certain planes.

Using *continuous radiation*, the diffuse spots are of course drawn out into radial streaks; the non-radial diffuse streaks and areas of strong diffuse scattering are extended radially and contribute to form a non-uniform background, and superimposed on these are the usual white radiation selectively-reflected streaks (for rotating or oscillating crystals) or Laue spots (for stationary crystals).

Since the intensity of diffuse scattering is, at most, only a few per cent of the intensity of selective reflection (and is generally much less), considerable x-ray exposures are necessary for satisfactory photographic observation, and most experimenters have been content to use the *direct characteristic radiation*, filtered or unfiltered, from a Ag, Mo, Cu or Fe target, together with its accompanying white radiation, thus obtaining photographs which combine the effects due to both types of primary radiation, but with an intensity which depends upon their relative predominance in the primary beam. Thus a very well-exposed Laue photograph of a stationary crystal, taken with direct characteristic radiation, will usually show both Laue and diffuse spots (plates 1*b*, 2*b*, 3*b*). The former occur at the usual $2\theta_L$ angles of deflection corresponding to wave-lengths in the continuous part of the primary beam, while the latter occur at angles of deflection that are *near* to $2\theta_B = 2\sin^{-1}(\lambda/2d)$, where λ is the characteristic wave-length and d is the spacing of a set of planes which is *nearly* inclined at θ_B to the incident rays. The presence of Laue spots has the advantage of enabling the crystal orientation to be determined with precision; it has the disadvantage of sometimes obscuring details of the diffuse background.

No experiments have yet been carried out specifically to determine the wave-length of the Faxén-Waller scattered radiation. There is, however, no evidence of an increase of appreciably modified radiation in the neighbourhood of Bragg reflections, and theory predicts that the wave-length of the scattered radiation, although slightly different from that of the incident radiation, will not be *measurably* changed. The frequency of the atomic vibrations is, at the most, about $10^{13}\,\text{sec}^{-1}$, whilst that of x radiation is of the order of $10^{18}\,\text{sec}^{-1}$. The change of frequency of the scattered radiation from 10^{18} to $(10^{18} \pm 10^{13})$ will only modify the scattered wave-length by 0·001%, an amount which is just about at measurable limit under the most favourable conditions. Although the frequency change has been emphasized both by v. Laue (1941) and by Raman (Symposium. 1941, 42), it is really somewhat misleading to call the strong spots due to thermal scattering "modified reflections", because this would imply a change of wave-length which, if not actually measured was at least measurable. The expressions "diffuse scattering" or "diffuse reflections", which are purely descriptive, are much better (Zachariasen, 1941 b). The question of the coherence of the diffuse scattering is also one that cannot be settled by direct experiment; but it is clear that if, as theory indicates, the diffuse reflections are due to partial reinforcement of radiation scattered by electrons belonging to different atoms whose vibrations, though not independent, are not exactly in phase, then the contributing scattered rays will be *incoherent* (Born and Sarginson, 1941; v. Laue, 1941. See also Jauncey and Compton, 1927, for a discussion of the coherence of Bragg reflections).

Plate 1 *a*.

Stationary crystal of Na Cl, monochromatic Mo $K\alpha$ radiation along cube-axis direction ; 150-hour exposure, room temperature.

(*Reproduced by kind permission of the Editors of the ' Physical Review '.*)

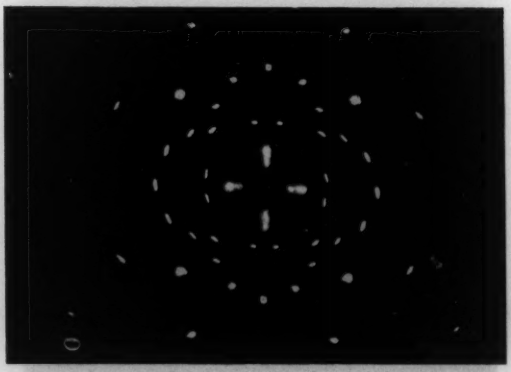


Plate 1 *b*.

Stationary crystal of Na Cl, direct radiation from Mo target along cube-axis direction; $2\frac{1}{2}$-hour exposure, room temperature.

(*Reproduced by kind permission of the Editors of the ' Physical Review '.*)

To face page 208

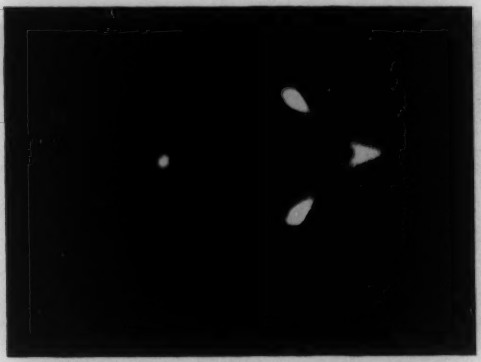


Plate 2 *a*.

Stationary crystal of Na, monochromatic Cu $K\alpha$ radiation at $(\theta_B - 2°) = 19°{\cdot}3$ to (002) plane, [100] axis vertical. The strong diffuse reflections are due to the regions 002, $0\bar{1}\bar{1}$, $10\bar{1}$ and $\bar{1}01$. The inner circle is due to the glass tube containing the Na specimen. Exposure 1 hour; room temperature.

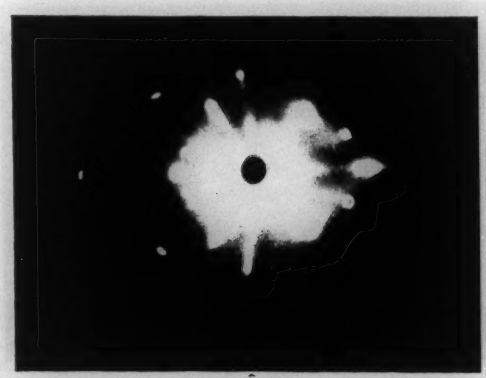


Plate 2 *b*.

Stationary crystal of Na, direct radiation from Cu target, same orientation as above. Note strong diffuse streaks due to the continuous radiation, in addition to the characteristic diffuse reflections. Exposure 3 minutes; room temperature.

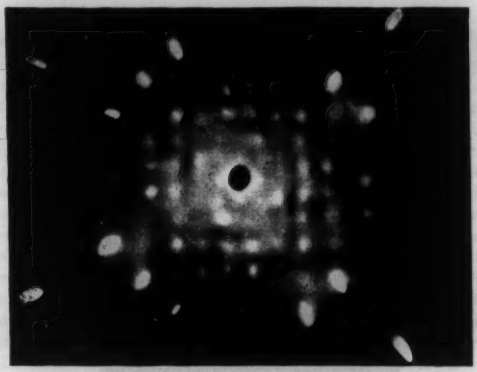


Plate 3 *a*.

Stationary organic crystal (pentaerythritol tetraphenylether) ; monochromatic Cu $K\alpha$ radiation along tetragonal axis. Exposure 1½ hours ; room temperature.

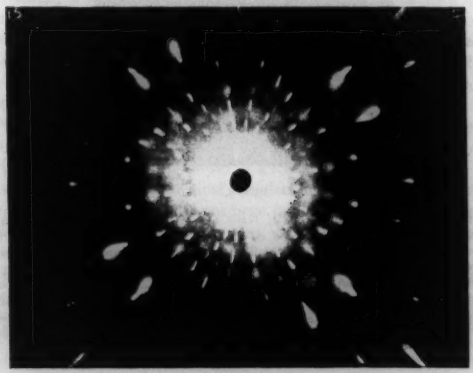


Plate 3 *b*.

Same as above, but using direct radiation from Cu target. Exposure 15 minutes ; room temperature.

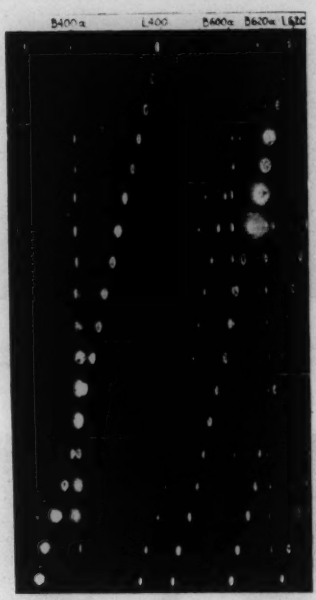


Plate 4.

Equatorial reflections from cleavage face of NaCl of the direct radiation from a Cu target. Glancing angle of incidence θ_i is 45° 7′ ($=\theta_B + 12°$) for top strip, decreasing by 1° for each successive strip. The Bragg (Cu $K\alpha$) reflections 400, 600, 620 are placed as fiducial marks on each strip ; the Laue reflections form slanting lines on the set of strips ; the diffuse 620 reflections are seen on the upper strips, and the diffuse 400 reflections on all strips from $\theta_i = 40\ 7' \rightarrow 28°\ 7'$. It may easily be seen that there is a displacement of the diffuse reflection relative to the Bragg reflection, in the direction of the Laue reflection, and that the diffuse reflection is most intense when it coincides with the Bragg and Laue reflections.

(Reproduced by kind permission of the Editors of the 'Physical Review'.)

(ii) *Ionization spectrometer observations*

(*a*) *Laval's experiments*

The Debye theory of thermal scattering implied that apart from the Laue interference peaks the scattering from a crystal would be independent of the crystal orientation. This was contradicted by the later Faxén-Waller theory, but none of the extensive and careful experiments on diffuse scattering carried out by Jauncey and his collaborators really tested this point properly, because in all of them the technique adopted avoided the positions of orientation—sensitive scattered radiation (Jauncey and May, 1924; Jauncey and Claus, 1934; Jauncey and Bruce, 1936; these are typical of many other papers). These experimenters noticed that there was a shoulder to each Laue reflection that was rather wider than the geometry of the slits would explain, but they were careful to make their measurements only in regions where the background was uniform.

The first experiments that were carried out definitely to test the Faxén-Waller theory were those of Laval (1938, 39), and these proved conclusively, by means of ionization-spectrometer methods, that when monochromatic radiation is used there exist regions of fairly well-defined strong diffuse scattering on a background of weak diffuse scattering. Laval used Mo $K\alpha$ and Cu $K\alpha$ radiation, monochromatized by reflection from a curved quartz crystal. The x rays were scattered from the surfaces of large carefully selected crystals of sylvine (cleaved, or cut in various ways and carefully polished), rocksalt (cleaved), aluminium (electrolytically polished), calcite (cleaved) and one perfectly clear diamond, and from a powdered specimen of silver whose micro-crystals were less than 10^{-12} cc. in volume; they then entered an ionization chamber filled with argon, by means of which Laval could detect the arrival of one photon of Mo $K\alpha$ every two seconds. By placing a suitable condenser in parallel with his quadrant electrometer Laval could also measure the incident beam intensity and thus reduce his results to an absolute scale. The incident beam produced an ionization current in the argon of 2.10^{-8}–10^{-10} amps.; the diffusely scattered beam one of 10^{-13}-8 10^{-16} amps., the latter being measurable to 10%. He expressed the "diffusing power" P of his crystals as the ratio of the scattering power per electron to the Thomson scattering power of a free electron.

Laval's work, which includes the most complete investigation of the scattering from one substance (sylvine) that has yet been published, appeared in a journal that is not readily accessible to most physicists, and his results will, therefore, be given here in some detail.

In a preliminary experiment he set the crystal face which was to be the scattering plane normal to the bisector of the directions of the incident and scattered beams; keeping these fixed, he turned the crystal face in its own plane about this bisector, thus changing the azimuth σ of the scattering plane. He found that for both strong and weak diffuse scattering P was independent of σ. Later experiments by Lonsdale and Smith (1941 b, 41 c, 42 a) confirmed this for sylvine, but showed that it is not true for all crystals.

In a more detailed study, Laval explored several regions of strong scattering

and of the weak background scattering, expressing the orientation of the incident and reflected beams relative to his crystal framework, in terms of the reciprocal lattice.

Let OX, OY, OZ be the reciprocal lattice axes, O the origin, $\overrightarrow{IO}$ the direction of the incident beam, $\overrightarrow{ID}$ the direction in which the scattering is to be investigated. Then if $IO = ID = 1/\lambda$, D, which Laval calls the pole of diffusion, will lie on the sphere of reflection corresponding to the primary wave-length λ (figure 1), and by suitably altering the directions $\overrightarrow{IO}$, $\overrightarrow{ID}$, D can be made *any*

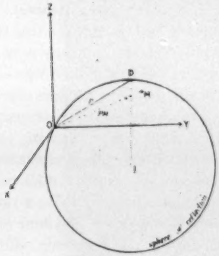


Figure 1.

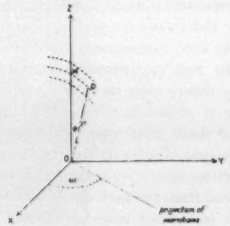


Figure 2.

point in reciprocal space within the limiting sphere, including the reciprocal lattice points, one of which is shown at M. By determining the value of P, the scattering power, as D moved in the neighbourhood of M, Laval mapped out the "domain" of strong scattering power about several reciprocal lattice points (i.e. about several Bragg reflecting positions). He expressed the position of D in the reciprocal lattice by means of the polar coordinates ρ, ψ, ω. Here $\rho = OD = \dfrac{2\sin\chi/2}{\lambda}$, where χ is the total angle of scattering OID*; therefore, $\rho_M = 1/d$, where d is the spacing of the planes corresponding to the reciprocal lattice point M.† The colatitude ψ is the angle ZOD, and the longitude ω is the angle between planes ZOX and ZOD. The domains which he particularly investigated for sylvine were 002, 004, 006, 022 and 044; he therefore chose his reciprocal lattice axes so that OX was along [100] and OZ along [001] or [011]. In each series of experiments, therefore, M was on the z axis and ψ measured the angular separation between $\overrightarrow{OM}$ and $\overrightarrow{OD}$.

Variation of ρ with crystal orientation. Keeping $\omega = 45°$, relative to [100] as OX (i.e. keeping D in the plane of [110] and [001]) and changing ρ by definite amounts, Laval examined the variation of P with changing ψ (figure 2). His

* Laval used 2ϕ for the total angle of scattering; this has been changed to χ to conform with notation used throughout this *Report*.

† In some of Laval's tables he expresses ρ in terms of $\rho_{100} = 0{\cdot}159_8$ (KCl) as unit. Here ρ will be given throughout in terms of $(\text{A.})^{-1}$ or, more strictly, $(\text{kx.})^{-1}$

results are shown in figure 3. Only one-half of each curve is shown (except for $\rho=\rho_M$), the curves being symmetrical about the x axis, but not about the point M. Curves for $\rho>1/d$ are shown to the right, those for $\rho<1/d$ to the left of the P axis in figure 3. The angle of mis-setting (grazing angle of incidence θ_i—Bragg angle θ_B) is given for each curve, and also the corresponding values of ρ/ρ_M. It will be seen from the curves that for each value of ψ, P has a maximum value (Γ) when D is nearest to M, and decreases to a steady value (γ) between reciprocal lattice points. It will be seen later that γ is about one-half of the value of the

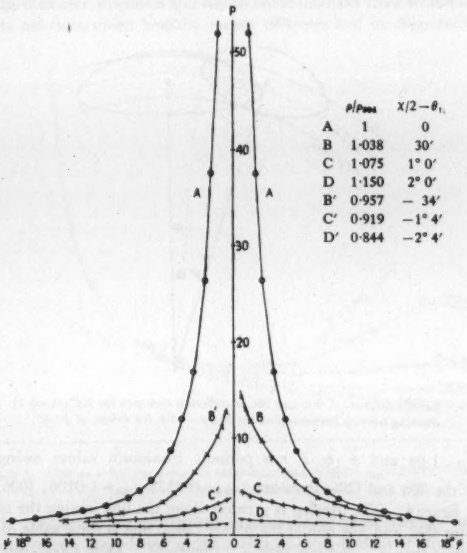


Figure 3. Laval's measurements of P for KCl 004, $T=290°$ K., $\omega=45''$, $\lambda=0{\cdot}71$ A

average unmodified background scattering, as measured by Jauncey and others. The value of P for points very near to M could not be measured because of the finite size of the Bragg reflection. One important observation made by Laval was that P was greater when $(\rho-\rho_M)$ was positive than when it had an equal negative value. This is in agreement with the Faxén-Waller theory, according to which P varies as ρ.

Then keeping ρ, ψ constant, Laval investigated the influence of changing ω. He found that for KCl, and for positions of D near to M, the scattering power P

was independent of ω; that is to say, the domains of strong scattering are regions of revolution about OM. In the case of 002 this is true for all values of ρ and ψ, the domain being entirely surrounded by a region of weak scattering. For the 004 domain, if $0{\cdot}60<\rho<0{\cdot}70$ and if $\psi>18°$, P is periodic, having a maximum value for $\omega=0, \frac{\pi}{2}, \pi, \frac{3\pi}{2}$ and a minimum value for $\omega=\frac{\pi}{4}, \frac{3\pi}{4}, \frac{5\pi}{4}, \frac{7\pi}{4}$. This Laval explained as being due to the fact that along the directions [100], [010], [$\bar{1}$00] and [0$\bar{1}$0], the 004 and {204} domains overlap. Similarly

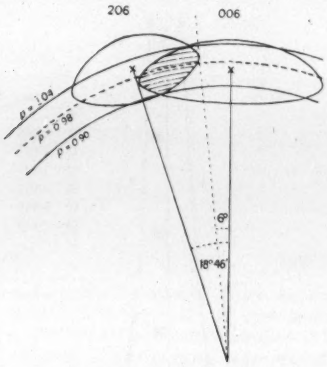


Figure 4. Laval's diagram of 006 and 206 isodiffusion contours for KCl, $\lambda=0{\cdot}71$, $\omega=0°$ showing overlap between $\rho=0{\cdot}90$ and $\rho=1{\cdot}04$, for values of $\psi>6°$.

if $0{\cdot}90<\rho<1{\cdot}04$ and $\psi>6$, P has periodic maximum values owing to the overlap of the 006 and {206} domains ($\rho_{006}=0{\cdot}957$; $\rho_{206}=1{\cdot}0106$; $[006]\widehat{:}[206]=18°26$; figure 4). The overlap is greatest along the line joining the reciprocal lattice point 006 and any one of the points {206}, that is for a value of $\rho=0{\cdot}98$ rather greater than ρ_{006} and less than ρ_{06} (figure 5). Thus Laval obtained the first definite proof that *along certain directions the domains stretch right across the boundaries of the Brillouin zones*. The minimum value of P along the 006–206 bridge of strong scattering is about 3·22, as compared with 1·39 for the real "weak scattering" γ, for the same value of ρ. The Compton scattering (0·67) is included in both these figures.

For a given value of ω, then, the minimum P found as ψ changes is not necessarily the absolute minimum for that domain. It is the absolute minimum γ (for the cases above considered) only if $\omega=(2n-1)\frac{\pi}{4}$. The maximum value of P

cannot be measured because of the finite size of the Bragg reflection, but if D is not too near to M, the maximum value Γ of P along any meridian can be estimated by a reasonably accurate extrapolation, and $(\Gamma-\gamma)$ can be taken as a *measure of the effect of crystal orientation* for any value of ρ. The boundaries of the domain occur for $\Gamma-\gamma=0$, but as this imposes a condition difficult to measure exactly, Laval has stated that $\rho_{max.}$ and $\rho_{min.}$ are arbitrarily taken as those values of ρ for which $\Gamma_2-\gamma=\gamma/10$ and $\Gamma_1-\gamma_1=\gamma_1/10$, γ_2 and γ_1 being the weak scattering corresponding to $\rho_{max.}$ and $\rho_{min.}$

In this way the topography of the domains has been determined in a few cases; the diameter, thickness and rate of fall of intensity from within the domain outwards has been given both by means of tables and in diagrammatic form,

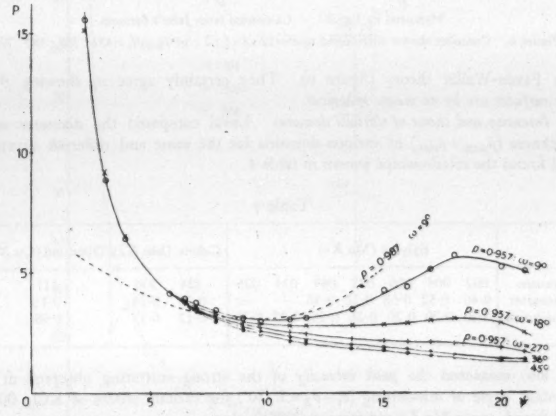
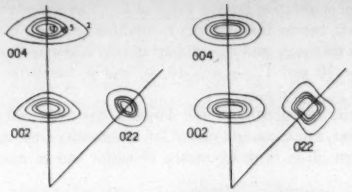


Figure 5. Diffuse reflecting power in neighbourhood of 006 : $\rho_{006}=0{\cdot}957$; $T=290°$ K., $\lambda=0{\cdot}71$ A., as measured by Laval.

the only completely uninvestigated regions being those already occupied by Bragg reflections, these being at the centre of each domain of strong scattering.

Isodiffusion contours. Laval gives some of his results for KCl in the form of isodiffusion curves in reciprocal space (curves of constant P); these are plotted for 002, 004 and 022, but from the data given in the tables it does not appear that his knowledge of the shape of the contours is quite so exact as the diagrams imply. In the case of 022, for example, the values of P for $\rho=\rho_{022}$ and $\psi=1°\ 34'\rightarrow20°\ 2'$ are given, and it is stated that these are constant for all values of ω. Values of $(\Gamma-\gamma)$ are also given for various values of $(\rho-\rho_{022})$. This, in effect, does no more than state the scattering power on the cap of a sphere having OM as its radius and M as the centre of the cap, and along a portion of OM produced.

On the whole, however, Laval's isodiffusion contours or surfaces agree reasonably well with those calculated for KCl according to Jahn's extension of



Measured by Laval. Calculated from Jahn's formula.

Figure 6. Contours shown correspond to $P=12:8:5:2$; or $I_{KCl}/I_e=432:288:180:72$.

the Faxén-Waller theory (figure 6). They certainly agree in showing that *the surfaces are by no means spherical.*

Intensity and shape of various domains. Laval compared the diameter and thickness ($\rho_{max.}-\rho_{min.}$) of various domains for the same and different crystals and found the relationships shown in table 4.

Table 4

	Sylvine (Mo $K\alpha$)							Calcite (Mo $K\alpha$)		Diamond (Cu $K\alpha$)
Domain	002	004	006	022	044	024	026	224	336	111
Diameter	0·40	0·52	0·68	0·32	0·46	—	—	0·20	0·26	0·15
Thickness	0·20	0·20	0·20	0·26	0·26	0·27	0·30	0·13	0·13	0·08

He also measured the *peak intensity* of the strong scattering observed at a constant angle of mis-setting ($\theta_i-\theta_B=2°\ 30'$) for various orders of KCl, 001, and found, using Mo $K\alpha$ radiation at 290° K., that

$$002:004:006:008:0010=100:50:33:21:14.$$

He is careful to state, however, that in making these observations he used wide-open slits, thus favouring the more intense maxima at the expense of the weaker. His curves also indicate an increase in half-width at half-maximum for the higher orders as compared with the lower (see figures for "diameter" given in table 4), and the conclusion appears to be that the *total intensity* of the strong scattering for any position of mis-setting does not fall off markedly with order, certainly not to anything like the extent found for the integrated Bragg reflections, for which James and Brindley (1928) give

$$002:004:006:008:0010=100:21{\cdot}8:5{\cdot}31:2{\cdot}27:0{\cdot}40.$$

An even more striking comparison of the strong scattering for various orders of the same plane may be obtained by plotting P against ($\rho\psi$) from Laval's tables

of $P:\psi$ for KCl 002, 004 and 006. These were measured with narrow slits, and it is found that the points lie almost exactly on one and the same curve, except for large values of $\alpha\psi$, for which P is slightly higher for the higher orders, corresponding to a gradual increase of the weak scattering as the total angle of deflection χ increases (figure 7). The apparent increase in diameter observed by Laval is due simply to the fact that for 002 P reaches a steady minimum value rather more quickly than for 004 or 006; for the higher orders, P continues to decrease, very slowly, over a larger angular range. The actual domains

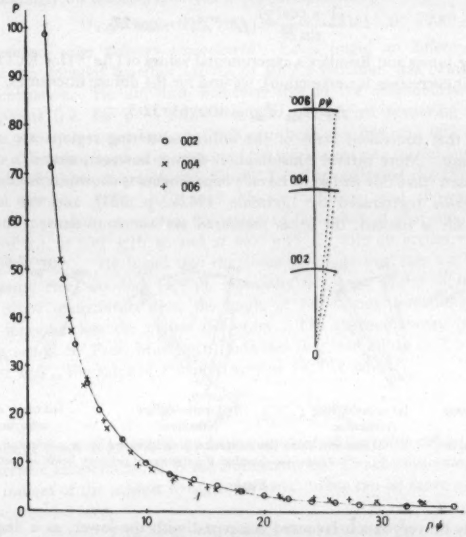


Figure 7. Absolute intensity of diffuse reflecting power for the 002, 004 and 006 regions of KCl ; $T=290°$ K. ; $\lambda=0{\cdot}71$ A. ; as measured by Laval.

are, in fact, ***almost identical both in regard to absolute intensity and intensity distribution***.

It is difficult to believe that this identity is not, to a certain extent, fortuitous. The theoretical expression for the intensity of diffuse scattering at any point certainly indicates that the falling off of intensity with scattering angle, if any, will be much less than for the Bragg reflections. The expression for the diffuse intensity will contain the usual polarization factor $(1+\cos^2 2\theta)$, or more strictly $(1+\cos^2\chi)$, the atom-scattering factor f_0 and its temperature modification e^{-M}, the variation of which with scattering angle is the same for both Bragg and

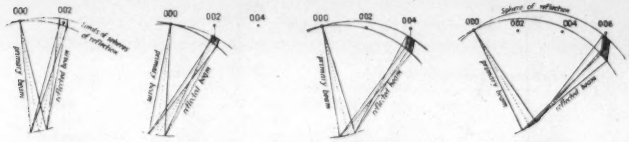
diffuse reflections. It probably includes also the usual factor $\frac{1}{\sin 2\theta}$, if the intensity of diffuse scattering from a small crystal or an extended crystal face is measured using a divergent beam. In addition, however, it includes a factor $(1 - \cos 2\theta)$, which is quite foreign to the Bragg intensity and which arises essentially from the fact that small displacements will affect the high-order intensities much more than the low orders. Introducing all these terms, we have

$$I \propto \frac{(1 + \cos^2 \chi)}{\sin 2\theta} (f_0 e^{-M})^2 (1 - \cos 2\theta),$$

and, using James and Brindley's experimental values of $(f_0 e^{-M})$ for KCl at 290° K. (with their correction for extinction), we find for the diffuse intensities,

$$I_{00\cdot} : I_{004} : I_{006} = 100 : 63 : 32{\cdot}5,$$

provided that equivalent areas of the diffuse scattering regions are measured in each case. More careful consideration shows, however, that if a divergent beam is used (that this was so in Laval's experiments is shown in the diagram of his apparatus reproduced by Lonsdale, 1942c, p. 321), and the ionization chamber slit is limited, the areas measured are *not* equivalent. The higher



Integrated Bragg reflection. (Moving crystal or divergent beam.)

1st order diffuse reflection.

2nd order diffuse reflection.

3rd order diffuse reflection.

(Diagrams show the increasing area included by rays reflected into the ionization chamber if a divergent primary beam is used.)

Figure 8.

orders may be very much favoured compared with the lower, as a diagram in reciprocal space soon shows (figure 8). The value of this factor cannot be estimated unless the exact divergence of the primary beam (given by Laval as 40′), the dimensions of the ionization chamber slit and the distribution of scattering power in the diffuse regions are known.

At present, therefore, it does not seem possible to indicate precisely what variation with scattering angle would be theoretically expected, although clearly there will be no marked falling-off. For orders of reflection high enough for the amplitudes of vibration to be no longer small compared with the effective spacing, the thermal theory, as at present formulated, is not really applicable. Both theory and experiment need to be extended along these lines.

Dependence upon substance. Laval found that the intensity of strong diffuse scattering from NaCl 004 is less than that from KCl 004, although the former

shows the greater selective reflection (James and Brindley give NaCl, 004, $98{\cdot}41 \times 10^{-6}$: KCl, 004, $60{\cdot}0 \times 10^{-6}$). He also showed that the cleavage planes of calcite, which have approximately the same atomic density as those of sylvine (002), gave much less diffuse scattering, especially in high orders, while the scattering from diamond was the smallest found from any crystal. The order $P_{\mathrm{K\,Cl}} > P_{\mathrm{Na\,Cl}} > P_{\mathrm{calcite}} > P_{\mathrm{diamond}}$ (conditions being matched) is, according to the thermal theory, a consequence of the increasing order of the elastic constants or, correspondingly, of the characteristic Debye temperatures:

$$\Theta_{\mathrm{K\,Cl}} : \Theta_{\mathrm{Na\,Cl}} : \Theta_{\mathrm{calc.}} : \Theta_{\mathrm{diam.}} = 230^\circ : 281^\circ : 520^\circ : 1900^\circ \text{ K}.$$

Dependence upon primary wave-length. Laval found no difference in the shape of the domains whether Mo $K\alpha$ or Cu $K\alpha$ radiation was being used for the investigation. He did find, however, that in the neighbourhood of the lattice points (i.e. for $\psi < 3^\circ$) the absolute intensity of scattering of Cu $K\alpha$ radiation was up to twice that of Mo $K\alpha$ for both the 004 and 002 reflections of K Cl. The same applied in respect of the weak scattering for values of $\rho < 0{\cdot}3$.

Dependence upon temperature. Finally Laval investigated the effect of temperature change. Using KCl, with Mo $K\alpha$ radiation, and allowing for the change of spacing on expansion, he compared the values of P at room temperature (290° K.), at $550^\circ \pm 10^\circ$ K. and at $665^\circ \pm 15^\circ$ K., with an accuracy which he estimates at 10%. He found that the domain volume increases markedly with temperature rise; so does $(\Gamma - \gamma)$, especially for large angles of mis-setting. For a given temperature rise, the angle of mis-setting remaining constant, $(\Gamma - \gamma)$ increases less the higher the order. The thermal theory (see accompanying article by Prof. Max Born) indicates that, for values of $T > \Theta$ and not too close to T_h, the value of P should vary as Te^{-T/T_h}, where

$$T_h = \frac{mka^2\Theta^2}{3\hbar^2} \cdot \frac{1}{h_1^2 + h_2^2 + h_3^2} \qquad \ldots\ldots(7)$$

(m is atomic mass in gm., a is the lattice constant of the cubic cell in cm., Θ the characteristic temperature, k, h the Boltzmann and Planck constants, $h_1 h_2 h_3$ the Miller indices of the nearest Bragg reflection). This can be more conveniently expressed in the form

$$T_h(h_1^2 + h_2^2 + h_3^2) = \mu A^2\Theta^2 \cdot 1{\cdot}759 \times 10^{-4}, \qquad \ldots\ldots(8)$$

where μ is atomic mass in atomic units. A the lattice constant in angstrom units. For KCl, $\mu = (35{\cdot}5 + 39{\cdot}1)/2 = 37{\cdot}3$, $A = 6{\cdot}28$, $\Theta = 230^\circ$ K.; hence $T_h(h_1^2 + h_2^2 + h_3^2) = 13670^\circ$. Laval particularly investigated the 002 and 004 domains for $\chi = 2\theta_{002} = 12^\circ\ 58'$, $\omega = 45^\circ$ and $\psi = 2^\circ \rightarrow 11^\circ$ and for $\chi = 2\theta_{004} = 26^\circ\ 6'$, $\omega = 45^\circ$ and $\psi = 1^\circ\ 25' \rightarrow 16^\circ\ 45'$, at the three temperatures given above. In table 5 his results are given, first on an absolute scale, then with the high temperature measurements expressed as a percentage of those at room temperatures, and finally his mean results are compared with those calculated from the expression

$$P \propto Te^{-T(h_1^2 + h_2^2 + h_3^2)/13670}.$$

(Born and Lonsdale, 1942). The agreement is well within the limits of experimental error (10%), and, moreover, the comparative constancy of P_T/P_{290}° for

Table 5

ψ	P_{002} (absolute) 290°	550°	665° K.	P_{002} (relative to 100 at 290°) 550°	665° K.	ψ	P_{004} (absolute) 290°	550°	665° K.	P_{004} (relative to 100 at 290°) 550°	665° K.
°						° ′					
2	72	114	132	158	183	1 25	51	73·5	75	144	147
3	48	77	89	160	185	1 50	38·5	54	56·5	140	147
5	26·4	44	53	167	201	2 20	26·5	38	40	143	151
6	20·5	33	41	161	200	3 20	17	23·6	25	139	147
7	17·7	28	35	160	200	4 20	12	16·8	18	140	150
8	15	25	32	167	213	5 15	8·9	12	13·5	135	152
9	12·3	20	24·5	163	199	6 15	***7·5**	**10**	**12·3**	**133**	**164**
10	***9·0**	**16·5**	**20·2**	**183**	**224**	6 55	5·1	6·8	7·7	133	151
11	8·5	14	17	165	200	7 55	4·4	5·8	6·5	132	148
						8 50	3·7	4·8	5·5	130	149
						10 45	2·95	3·85	4·5	130	152
						12 45	2·5	3·28	3·85	131	154
						14 45	2·1	2·75	3·2	131	152
						16 45	1·7	2·3	2·65	135	156
Mean observed %				165	201	Mean observed %				135·5	151
Calculated %				176	207	Calculated %				140	148

* These figures are inconsistent with those given elsewhere by Laval. The percentages given are therefore doubtful.

differen· values of ψ (corresponding to a range of elastic ·wave-length 90→16·5 A. for 002 and 127→10·7 A. for 004) shows that it is legitimate to assume that the intensity distribution throughout the domain does not markedly vary with temperature. Laval's statement that the meridians became somewhat deformed as the temperature rose is, in fact, hardly confirmed by his own tables of results, since (apart from one doubtful measurement) the variation in P_T/P_{290} is within the limits of his estimated experimental error.

Laval himself made no numerical comparisons between the results of his observations and detailed calculations from the thermal theory, but from a purely general consideration of the experimental facts as he found them he oncludes: "Il n'est pas douteux que la diffusion forte soit due aux ondes élastiques produites par l'agitation thermique modulant le réseau crystallin de telle sorte que chaque nœud du réseau polaire s'étale suivant un domaine entier de la diffusion forte. . . . . On peut espérer que la topographie des domaines de la diffusion forte apportera des renseignements précis sur la distribution de l'énergie entre les diverses ondes élastiques qui constituent l'agitation thermique".

(*b*) *Baltzer's experiment*

Later ionization spectrometer measurements were made by Baltzer (1941) on rocksalt, at 300° and 100° K., using filtered Cu $K\alpha$ radiation with a weak background of continuous radiation. He also found a variation of intensity change with order of reflection (table 6).

Table 6

Region of reflection	200	400	600
Diffuse scattering at 300° K. / Diffuse scattering at 100° K.	2·25	2·18	1·91
Angle of mis-setting $(\theta_i - \theta_B)$	$+3°$	$+3°$	$-4°$

At temperatures low compared with the Debye temperature (281° K. for NaCl) the relationship, diffuse scattering intensity $\sim T \cdot e^{-T/T_h}$, will not be expected to hold. Baltzer suggests that reasonably good agreement may be obtained using the relationship, intensity $\sim(1-e^{-2M})$, where M is the Debye-Waller temperature factor (formula (4)). For NaCl, a mean calculated value of $2M$ is

$$7{\cdot}78 \times 10^{-5} \cdot T(h_1^2+h_2^2+h_3^2)Q\left(\frac{\Theta}{T}\right),$$

from which

$$\frac{(1-e^{-2M})_{300\,\text{K.}}}{(1-e^{-2M})_{100\,\text{K.}}} = 2{\cdot}49 \text{ for (200)};\ \ 2{\cdot}30 \text{ for (400)};\ \ 2{\cdot}01 \text{ for (600)}.$$

Waller and James' (1927) observed values of M_{Na} and M_{Cl} lead to similar values, namely. 2·42, 2·23 and 1·99, which are in fair agreement with Baltzer's observations. Baltzer's measurements were not all made at the same angle of mis-setting, but he mentions that the intensity ratio for 300° : 100° K. did not vary appreciably for $2° < (\theta_i - \theta_B) < 7°$, the inference again being that the diffuse intensity distribution within each scattering region does not change appreciably with temperature.

(iii) *Photographic and photometric observations*

Other investigations, many of which were undertaken in ignorance of Laval's work, have confirmed or extended his observations. Owing mainly to the difficulties of communication during war-time, the experimental work since 1939 has been divided, along geographical lines, into three definite sections, workers in each of which have, generally speaking, related their results to three different formulations of theory. It will be simplest, therefore, to consider separately the experimental work done in the U.S.A., India and England, remembering that the investigations were carried on very nearly simultaneously.

American papers. In 1938 Wadlund found unexpected spots and radial streaks on Laue photographs of NaCl and KCl, using radiation from a tungsten target, but he offered no explanation of them. Zachariasen (1940) suggested that they were due to thermal agitation and put forward a simplified, but imprecise, formulation of the Faxén-Waller theory which was, however, much more practical than the original theory in that it included an estimate of intensity, position of maximum and value of half-width at half-maximum intensity for diffuse spots corresponding to any given crystal orientation, all in terms of measurable

quantities. These formulae were tested (Siegel and Zachariasen, 1940; Siegel, 1941; Jauncey and Baltzer, 1940, 41; Gregg and Gingrich, 1941; Kirkpatrick, 1941), but the agreement with experiment was not convincing. One of the assumptions made by Zachariasen—that of an equal velocity for all elastic waves in the crystal—was, however, obviously far too crude, and later (1941 c) he gave a more rigorous calculation of formulae for a simple cubic lattice of the Born-Kármán type (having $c_{12}=c_{44}$), and in particular for the two cases $c_{11}\approx 3c_{12}$ and $c_{11}\gg 3c_{12}$.* The experimental results of Jauncey and Baltzer (1941) were compared with this more rigorous theory and were found to agree within the limits of experimental error (table 7; Zachariasen, 1941 d; Jauncey, Baltzer and Miller, 1941).

Table 7. Observed and calculated data for NaCl

400			620			402		
$(\theta_i-\theta_B)$	$(2\theta_m-2\theta_B)$ obs.	calc.	$(\theta_i-\theta_B)$	$(2\theta_m-2\theta_B)$ obs.	calc.	$(\theta_i-\theta_B)$	$(2\theta_m-2\theta_B)$ obs.	calc.
°	°	°	°	°	°	°	°	°
7·0	2·4	2·7	3·87	4·5	4·4	1·82	0·75	0·85
6·0	2·4	2·3	2·87	3·2	3·2	1·02	0·5	0·5
5·0	2·1	1·9	1·87	1·8	2·1	0·20	0	0·1
4·0	1·6	1·5	0·87	0·9	1·0	−0·73	−0·15	−0·35
3·0	1·3	1·2	−0·13	0	−0·1	−1·35	−0·8	−0·8
2·0	0·8	0·8	−1·13	−1·2	−1·3	−2·42	−0·9	−1·2
1·0	0·6	0·4	−2·13	−2·3	−2·4	−3·33	−1·3	−1·6
0	0	0	−3·13	−3·6	−3·5			
−1·0	−0·4	−0·4	−4·13	−4·5	−4·7			
−2·0	−0·6	−0·8	−5·13	−5·4	−5·8			
−3·0	−1·2	−1·2						
−4·0	−1·8	−1·5						
−5·0	−1·8	−1·9						

In this table $(\theta_i-\theta_B)$ is the angle of mis-setting of the crystal from the Bragg position of the plane mentioned; $2\theta_m$ is the scattering angle corresponding to the diffuse maximum and $(2\theta_m-2\theta_B)$ is the *shift* of that maximum from the position of the Bragg reflection. In Jauncey and Baltzer's beautiful photographs (reproduced in plate 4) this shift was made obvious by using Cu radiation, scattered from the surface of a cleavage face of rocksalt (the shift for Cu $K\alpha$ radiation is several times as large as that for Mo $K\alpha$ radiation), and also by giving each film a double exposure, first with the crystal oscillating through a small angle about θ_B (exposure time less than 1 minute) and then with the crystal stationary at θ_i (exposure time 10 to 20 hours), so that the Bragg reflection, diffuse reflection and Laue reflection are all found on the same film. These photographs also show very clearly the essential diffuseness of the background spots as compared with the Bragg and Laue spots, and their decrease in intensity as the

* See the accompanying article by Professor Max Born for an outline derivation of Zachariasen's formulae. In Born's notation, $\theta_i-\theta_B=\phi$.

crystal orientation is varied away from the Bragg reflecting position. Another remarkable photograph (plate 1 *a*), taken by Gregg and Gingrich (1941) with monochromatic Mo $K\alpha$ radiation *through* a slip of rocksalt 0·6 mm. thick, with an exposure of 150 hours, shows diffuse maxima without any accompanying Laue spots, and these maxima are in some cases connected by streaks which pass right across the Brillouin zone boundaries. A second monochromatic photograph (not reproduced), taken with the crystal tilted through 11° from its former position, showed a pattern with the maxima in approximately the same positions as before, but with their relative intensities changed.

One of the most precise photometric studies of diffuse maxima shifts, intensity distributions, half-widths at half-maximum and integrated intensities has been carried out by J. H. Hall (1942) for various settings of KCl crystals, near to the 400, 420 and 440 Bragg reflections, at room temperatures only. She used filtered Cu radiation and crystal plates cleaved or ground and polished so that the half-width at half-maximum of the Bragg reflections was of the order of only 2′ of arc in each case. Considerable care was taken to eliminate air scattering, to correct for the vertical divergence (58′) of the incident beam (the horizontal divergence was only 9′ to 18′, and was therefore negligible), and to determine the relative effects of the Cu $K\alpha_1$ and α_2 components at the different angles of incidence used. Having done this, she found that the results were in good agreement with the predictions of Zachariasen's theory, taking into account the actual values of the elastic constants of KCl. She noticed a small difference in the intensities for positive and negative values of $(\theta_i - \theta_B)$, but used the average for purposes of comparison with theory. Her value of the ratio of the intensity of the diffuse maximum, at $\theta_i = \theta_B + 20'$, to the integrated intensity of the sum of α_1 and α_2 Bragg reflections, was of the order of 5×10^{-5}. She also obtained diffuse scattering data for K Br crystals at angles of mis-setting $\theta_i - \theta_{Bs_1} = -35'$ and $-18'$, and, by a trial and error method, found that the value of c_{11}/c_{12} which would give the shifts and half-widths observed, to the nearest degree of approximation, was $c_{11}/c_{12} = 3$. The agreement obtained was not, however, very good, and it may be added that this ratio does not correspond with the values of the elastic constants as measured by Bridgman (1929. See also Schmid and Boas, 1935, p. 21), namely. $c_{11} = 3{\cdot}33$, $c_{12} = 0{\cdot}58$, $c_{44} = 0{\cdot}62 \times 10^{11}$ dynes/cm^2.

P. Kirkpatrick (1941) and Teague (1941, 42) have drawn attention to the fact that crystal imperfections such as the existence of very small blocks or surface disorder can give extra non-Laue scattering which, if care is not taken, may be confused with the real effects due to thermal scattering.

Indian papers. Early in 1940, Raman and Nilakantan found extra reflections on Laue photographs of diamond, which they published as evidence of a supposedly new phenomenon. Their theory was that the primary x rays themselves excited characteristic lattice vibrations and were then reflected with a very small change of frequency (not measurable) by the new stratifications of density thus set up (Raman and Nilakantan, 1940, 41; Raman and Nath, 1940; Raman, 1941; Preston, 1942). This theory, which places the principal

emphasis on the high-frequency optical vibrations and which would lead to the rejection of the Born theory of lattice vibrations, has been criticized upon theoretical grounds (Sirkar and Gupta, 1940; Sirkar and Bishui, 1941; Born and Sarginson, 1941; Born, 1942), but from the point of view of this *Report* is to be judged by its measure of agreement with experiment.

Raman and Nilakantan (1941 a, 41 c) have pointed out that the extra reflections found for diamond are not diffuse but relatively sharp, and that they are not markedly temperature-sensitive. They suggest that this is a result of the excitation by the primary x rays of the well-defined 1332 cm^{-1} oscillation of the carbon atoms. Pisharoty (1941) has measured the absolute intensity of the 111 extra reflection from a single diamond crystal and has shown it to be relatively large—up to a few per cent of that of the Bragg reflection. He further states that, according to Raman's theory (Raman, 1941 b). the "forbidden" 222 diamond reflection is, in fact, an extra reflection of the same kind as the 111 extra reflection, and is not a genuine Bragg reflection at all, and that a similar, though weaker, extra 200 reflection should exist, but that there should be no extra reflection corresponding to the 220 Bragg reflection (Pisharoty and Subrahmanian, 1941; Raman, 1942). These are all facts that can be tested experimentally. They have been tested by Lonsdale and Smith (1941 a, 42 b; Lonsdale 1942 a, 42 c), who have proved that the extra 111 reflections referred to by Raman and Nilakantan, and Pisharoty, are *structure-sensitive*, a result which shows that they are not due to thermal vibrations, but that equally they are not explicable on Raman's theory. They are much more probably of the kind referred to by Kirkpatrick and Teague as being due to crystal fault or imperfection. Some rare diamonds (known as type II, whose peculiar transparency to infra-red and to ultra-violet radiation of certain frequencies has been studied by Robertson, Fox and Martin (1934, 36)) *do not show these extra* 111 *reflections at all*, although the [1332] vibration is just as characteristic of them as of type-I (normal) diamonds. *All* diamonds show a weaker and more diffuse kind of extra reflection (often partly masked by the strong extra reflections referred to above), which is just of the type to be expected on the Faxén-Waller theory, i.e. temperature-sensitive but not structure-sensitive. Lonsdale and Smith have called this the *primary* scattering and the sharp extra reflections the *secondary* scattering. Contrary to the predictions of the Raman theory, primary and secondary extra reflections of considerable intensity are found associated with the {220} Laue reflections, and also with the {113} and {331} reflections. Between the Bragg or Laue reflections and the extra reflections, whether of the primary or secondary type, there are marked experimental differences which enable them to be differentiated with comparative ease. For instance, if the crystal is a reasonably good one, Bragg reflection takes place only at a certain definite angle ($\pm$ a few minutes of arc), while the extra reflection persists over a range of angle of several degrees. Moreover, the Laue reflection shifts its position rapidly as the crystal rotates, while the extra reflection remains comparatively stationary. Lonsdale and Smith have shown, *using diamonds of both normal and rare types* (with Cu radiation), that the 222 "forbidden" reflection has

just the properties of a genuine, though weak, Bragg reflection and none of those of an extra reflection; they have also shown that no 200 reflection of any kind exists.

Raman and Nilakantan (1940b) have published Laue photographs of $NaNO_3$, NaCl and $CaCO_3$ showing many strong diffuse reflections, and have pointed out the remarkable resemblance between the pattern of diffuse spots and the layer-line diagram of an oscillating crystal. They have suggested equations for the determination of the positions of diffuse maxima (see also Bhagavantam and Bhimasenachar, 1940, 41; Nilakantan and Nayar, 1941), but although in some cases these are of geometrical value as showing the distribution of diffuse scattering power about the Bragg reflection, physically they are of doubtful significance (Zachariasen, 1941a; Jauncey, 1941; Lonsdale, 1941b). One new point established experimentally by Raman and Nilakantan is that the temperature changes of diffuse intensity and of Bragg intensity not only vary considerably from one substance to another, but that for non-cubic crystals they may vary differently from one plane to another of the same crystal, even for similar scattering angles. This, of course, is an extension to the diffuse scattering of the results referred to in §4 for hexagonal and other crystals. In the case of such crystals as $NaNO_3$, the possibility of the onset of group rotation at high temperatures, previously referred to, must always be borne in mind.

Venkateswaran (1941) has attempted to use measurements of the temperature variation of diffuse spot intensity of SiC, NaCl, $NaNO_3$ and $C(CH_2OH)_4$ for the calculation of the characteristic frequencies of those substances, but in doing so he has not allowed for the effect of temperature in modifying the atom-scattering factor. His experimental observations and their interpretation have been criticized by Lonsdale (1942b, 42c).

Many of the experiments carried out by Raman and his colleagues (Raman, 1941, 42; Raman and Nilakantan, 1941b; Pisharoty and Subrahmanian, 1941; Venkateswaran, 1941; Subrahmanian, 1941; Dayal, 1941) have been used in an attempt to disprove the thermal theory by showing that Zachariasen's original formulae did not explain the facts. This is now generally admitted; it does not appear, however, that any experimental facts have been given that are inconsistent with the more rigorous form of the thermal theory.

Photographs of organic crystals taken by Bhagavantam and Bhimasenachar (1941) and by Venkateswaran (1941) and Subrahmanian (1941) have shown effectively that the intensity of the diffuse reflections is closely related to the orientation of the molecules, a result which the thermal theory explains in terms of the variation of the elastic forces in the lattice.

English papers. In 1939 Preston published Laue photographs of single crystals of Al, NaCl and MgO, taken both with continuous and with monochromatized radiation from an Ag target, which showed diffuse spot patterns whose intensity was greatly enhanced by increasing the temperature to 500° c. He associated these patterns with the thermal theory of Faxén, which he believed to be thus proved to be correct. That similar diffuse spot patterns could be obtained with organic compounds was shown by Knaggs, Lonsdale, Müller and

Ubbelohde (1940), who published Laue photographs of benzil taken with radiation from a Cu target, the primary rays being at 0°, 10°, 15° from the crystal [0001] axis; they showed that the diffuse spot pattern was much less sensitive to crystal mis-setting than the ordinary Laue pattern. This work was followed up by Lonsdale, Knaggs and Smith (1940), who examined a number of inorganic and organic crystals, using Cu, Mo and Ag radiations, and who found that at the temperature of liquid air all diffuse spot patterns were weakened in intensity, sometimes to the point of disappearance, the Laue patterns becoming correspondingly clearer and more intense. The high temperature increases in diffuse scattering, found by Laval and Preston, were confirmed, and it was shown that all *these temperature changes were completely reversible* (Lonsdale, 1940).

A profusely illustrated experimental study of diffuse scattering patterns has been published by Lonsdale and Smith (1941 b). They took Laue, oscillation and rotation photographs, in many different orientations, of small single crystals of various typical inorganic and organic compounds, using Cu, Mo and Ag radiations, unfiltered, filtered or monochromatized. In order to avoid excessive exposures, they used a powerful 5-kw. x-ray tube (or in some cases a 50-kw. tube), by means of which a Laue photograph such as that shown in plate 3*b* could be obtained in 15 min. and a monochromatic photograph (plate 3*a*) in 90 min., without the use of intensifying screens. Rules are given in the paper for the indexing of diffuse spots and for the determination of the conditions under which such spots appear. The relations of the intensity, shape, size, position and persistence of the diffuse spots to the nature, structure, perfection, orientation and temperature of the crystal examined, to the radiation used, and to the conditions of photography, are considered in some detail. This survey included some monochromatic photographs of benzil taken with Cu $K\alpha$ radiation along and at right angles to the principal axis, which showed the remarkable resemblance of the diffuse spot pattern to the reciprocal lattice of the crystal, as projected on a plane normal to the incident beam. It was apparent that in this case the distribution of diffuse scattering power extended preferentially along the reciprocal lattice planes and axes, one notable exception being that little or no scattering occurred which would correspond to axes or planes passing through the origin. In terms of the thermal theory, the absence of such radial diffuse extension would indicate the absence, or small amplitude and high frequency, of waves of compression (longitudinal waves) as compared with those of the waves of shear (transverse waves). Non-radial streaks were found to extend continuously from one diffuse spot to the next, thus confirming Laval's observation that the reflecting power can extend in some cases right from one reciprocal lattice point to the next, and pointing to the existence of vibrations along certain directions of considerable amplitude even at high frequencies (Lonsdale, 1942 c). Lonsdale (1941 a) had already shown that layer structures gave one characteristically large, intense and persistent diffuse spot in association with the layer plane, and Lonsdale and Smith made a special study of such a reflection (for urea nitrate), showing by means of strip photographs how its

intensity and position varied as the crystal was turned through nearly 30°. They showed also that on photographs of chain structures there may be very long, thin, diffuse spots corresponding to planes which are normal, or nearly normal, to the chain axes. These different types of diffuse spot receive a simple and logical explanation in terms of elastic forces (Lonsdale, 1942c) and they may also be of considerable assistance as a subsidiary means of determining the orientation of the molecules in unknown structures (Lonsdale, Robertson and Woodward, 1941; Knaggs and Lonsdale, 1942).

Lonsdale and Smith (1941a, 41b, 42b; Lonsdale, 1942a, 42c; referred to on p. 282) have also carried out an investigation of the extra reflections given by diamonds of both types, using Fe, Cu and Mo radiations, and have interpreted their results in terms of the distribution of extra reflecting power in reciprocal space. They agree with Pisharoty (1940) in describing the *secondary* scattering associated with the {111} Bragg reflections from type I diamonds as due to sharp spike-like extensions of reflecting power along the six cube directions which intersect in the {111} lattice points; but they point out that in the case of the {220}, {113} and {331} reflections the extension of reflecting power is only along certain cube directions, not along all of them; while there is, apparently, no secondary scattering associated with the {400} Bragg reflections (though it is possible that an extra reflection may be masked by the 400 Laue reflection in all positions of the crystal). This incompleteness of the secondary scattering by type I diamonds, and its total absence in type II diamonds, has not yet received a satisfactory explanation.

Jahn (1941, 42a) has derived a simple theoretical expression* for the intensity of the scattering power I in a cubic reciprocal lattice, from a general consideration of the Faxén-Waller theory. I is given in terms of the radius vector and direction cosines ($\mathbf{r}$; l, m, n) of the point at which the scattering is being investigated, relative to the nearest reciprocal lattice point; the radius vector and direction cosines ($\mathbf{R}$; L, M, N) of the reciprocal lattice point relative to the origin; and the elastic constants (c_{11}, c_{12}, c_{44}) of the crystal.

$$I \propto \frac{\mathbf{R}^2}{\mathbf{r}^2}\{c_{44}{}^2+\Sigma[L^2\{c_{44}(c_{11}-c_{44})(m^2+n^2)+(c_{11}+c_{12})(c_{11}-c_{12}-2c_{44})m^2n^2\}]$$
$$-2\Sigma[MNmn(c_{12}+c_{44})\{c_{44}+(c_{11}-c_{12}-2c_{44})l^2\}]\}$$
$$\div\{c_{11}c_{44}{}^2+c_{44}(c_{11}+c_{12})(c_{11}-c_{12}-2c_{44})(l^2m^2+m^2n^2+n^2l^2)$$
$$+(c_{11}+2c_{12}+c_{44})(c_{11}-c_{12}-2c_{44})^3l^2m^2n^2\}. \qquad \ldots\ldots(9)$$

Jahn has pointed out that this involves a distribution of scattering power in reciprocal space which may vary considerably with the reciprocal lattice point considered, and has demonstrated this by means of stereographic projections of the isodiffusion surfaces for the main reciprocal lattice points of sodium (Jahn, 1942b), which he suggests as a particularly suitable substance for testing the formula. Lonsdale and Smith (1941c, 42a; Jahn and Lonsdale, 1942; Lonsdale, 1942c) have carried out this test, not only for sodium, but also for

* See Professor Born's article for the derivation of this formula; and compare also Sarginson, 1942.

lithium, lead and tungsten single crystals, and have obtained results in excellent agreement with Jahn's formula. The observations were made with radiation from a Cu target (plate 2 shows typical photographs of Na taken with unfiltered and with monochromatized radiation), on a succession of Laue photographs taken at small angular intervals in various crystal orientations, over ranges covering the observable regions of reflecting power for all possible planes. Na and Li single crystals are very soft and elastically anisotropic (for Na, $c_{11}=0{\cdot}52$, $c_{12}=0{\cdot}40$, $c_{44}=0{\cdot}41\times10^{11}$ dynes/cm^2; from Quimby and Siegel (1938), extrapolated to room temperatures). They give photographs showing intense non-radial streaks extending right across from one diffuse spot to the next. Jahn's formula only applies, strictly speaking, in the near neighbourhood of the reciprocal lattice points, since it takes no account of the high frequency optical vibrations, but it is clear that the diffuse spot patterns are due to vibrations covering the whole range of the frequency spectrum and that, properly interpreted, they should give information from which that spectrum can be determined. That the optical vibrations are usually, however, negligible by comparison with the low-frequency acoustical vibrations, is proved by the fact that Pb single crystals, which are elastically anisotropic in the same sense as Na (for Pb, $c_{11}=4{\cdot}77$, $c_{12}=4{\cdot}03$, $c_{44}=1{\cdot}44\times10^{11}$ dynes/cm^2; from Goens (1936), corrected for an arithmetical error), give photographs leading to similar, though less pronounced, reflecting regions, in spite of their being face-centred cubic in structure, whereas Na crystals are body-centred cubic. Tungsten crystallizes, like Na, with a body-centred cubic structure, but its diffuse spot pattern is completely different. Unlike Na, it is elastically isotropic and its elastic constants are large ($c_{11}=51{\cdot}3$, $c_{12}=20{\cdot}6$, $c_{44}=15{\cdot}3\times10^{11}$ dynes/cm^2; from Bridgman (1925) and Wright (1930)). For such elastic properties the Jahn formula predicts small, elliptical diffuse spots which should disappear at a very small mis-setting from the Bragg position; and this is exactly what Lonsdale and Smith have found.

The ionization-spectrometer observations of Laval (1939) on the 002, 004 and 022 reflecting regions of KCl (for which $c_{11}=3{\cdot}88$, $c_{12}=0{\cdot}64$, $c_{44}=0{\cdot}65\times10^{11}$ dynes/cm^2; from Försterling (1920), Bridgman (1929) and Durand (1936)) have given isodiffusion curves which, as already shown (figure 6), are in good agreement with those calculated from the Jahn formula. A photographic survey of KCl by Lonsdale and Smith (1941 b) has given similar agreement on a wider scale (Lonsdale, 1942 c; Jahn and Lonsdale, 1942).

To sum up: the static structure of the crystal decides at which reciprocal lattice points strong diffuse scattering can occur, but the precise shape, intensity and persistence of the diffuse pattern is controlled mainly by the quasi-elastic forces between the atoms. Detailed qualitative and some quantitative agreement with theory has been obtained for metals and for inorganic compounds having cubic structures, and the characteristic features of the patterns given by layer and chain structures are in accord with the general principles of the thermal theory; but on the one hand many interesting features of the diffuse spot patterns are not yet covered by theory, and on the other hand more precise quantitative

observations are needed for the exact confirmation of such formulae as have been already derived.

§6. DEBYE-JAUNCEY SCATTERING

As has already been mentioned in §3, an expression for the *average* "unmodified" diffuse scattering from an atom in a crystal is

$$\bar{I}_{\text{unmod. background}} = I_e(\Sigma f_{kk})^2(1-e^{-2M}). \qquad \ldots\ldots(10)$$

This is, in effect, the difference between the scattering of an atom when isolated from its neighbours (e.g. in the gaseous state) and the selective reflection in which the atom attempts to cooperate with its neighbours, a cooperation which is modified by the thermal motion of the atom about its mean position.

In a series of researches extending over many years, and embracing a variety of experimental methods, Jauncey and his colleagues have checked this formula and have even shown that by means of it not only could the atom-scattering factor Σf_{kk} and the temperature factor e^{-2M} be determined, but that the existence of zero-point energy could be proved, and that for the hexagonal crystal Zn the values of $M_{\parallel}$ and $M_{\perp}$ (or of $\Theta_{\parallel}$ and $\Theta_{\perp}$) could be determined and the anisotropy of the electron distribution in the atom detected with an accuracy at least equal to that given by methods depending upon observation of the Bragg reflections (Jauncey and Bruce, 1936, 37a; Bruce and McNatt, 1937; Baltzer and McNatt, 1939; McNatt and Jauncey, 1939; Jauncey and McNatt, 1939). In fact it is claimed (Jauncey and Bruce, 1937b) that whereas measurements on the temperature variation of Bragg reflections give only *mean* values of $\Theta_{\parallel}$ and $\Theta_{\perp}$ (parallel and perpendicular to the c-axis), diffuse scattering methods give values appropriate to each temperature at which observations are made. The f values obtained by such diffuse scattering experiments have also been used with fair success to check Hönl's dispersion formula (Bruce, 1938; McNatt, 1939; Jauncey, 1939).

The questions now arise: how is it that in the course of this careful research the relatively intense Faxén-Waller peaks were missed, and since these regions exist where the diffuse scattering is *greater* than the average, are there any well-defined regions where the diffuse scattering is *less* than the average, and if so, where are they?

The first of these questions is easily answered. The Faxén-Waller scattering is only observed as a distinct intensity *peak* when characteristic radiation is used (preferably with little or no continuous background) in the form of a rather narrow incident beam, the single crystal *and* ionization chamber or photographic plate being in such positions that the Bragg conditions for observation of selective reflection are nearly fulfilled. The simultaneous fulfilment of all these conditions never took place in any of the experiments carried out to test the formulae given by Debye, Jauncey and Woo (Compton and Allison, 1935). In order to increase the rather weak intensity of the diffuse scattering to be observed, fairly wide incident beams and a range of wave-lengths from the continuous spectrum of a tungsten tube run at high kv. were frequently used. Many observations were made on powder briquettes. Even when single crystals and monochromatic radiation were employed, the ionization method was used, the

chamber being set at a definite angle and the crystal orientation varied, or the crystal set at a fixed angle and the chamber position varied, *great care being taken always to avoid the exact or even approximate fulfilment of the Bragg conditions* (Jauncey and Williams, 1933; Jauncey and Claus, 1934; Jauncey and Deming, 1935b; Bruce and McNatt, 1937; McNatt, 1939). It has been suggested that rather perfect crystals are also essential for the observation of the Faxén-Waller scattering (Jauncey and Harvey, 1940), but the general surveys carried out by Lonsdale and Smith (1941, 42), who used very imperfect organic crystals as well as graphite, diamond, metals and inorganic crystals, have shown that this is not necessary.

With regard to the second and third questions: Jauncey and Harvey (1940) have suggested that it is more convenient to consider the "unmodified" diffuse scattering as being made up of two parts, a uniformly varying background of the Debye-Jauncey type, which is a function of $(\sin\chi/2)/\lambda$ alone, together with a positive *or negative* scattering which is, in addition, sensitive to the crystal orientation, but which is zero if averaged over a sufficient range of orientations of the crystal. The positive scattering, they suggest, accounts for the observed peaks or regions of strong scattering. They make no suggestion as to how the negative contribution is distributed. It seems clear from a study of the experimental data given by Laval (1939) that in that part of the reciprocal lattice near to the origin (that is, in the region of low-order reflection) the regions of strong scattering are isolated and are surrounded by regions of weak scattering which is *much less* than the Debye-Jauncey scattering (cf. figure 6 and table 1). As $(\sin\chi/2)/\lambda$ increases, the regions of strong scattering become larger and interpenetration takes place, while the regions of weak scattering become, as it were, lakes connected by canals. These regions of weak scattering have entirely disappeared in the case of calcite for values of $(\sin\chi/2)/\lambda > 0{\cdot}75$. A comparison of Laval's measurements of the weak scattering for KCl with that calculated from the Debye-Jauncey theory is shown in figure 9. It will be seen that the curves approach each other in the neighbourhood of $(\sin\chi/2)/\lambda = 2$, but that at room temperatures the weak diffuse scattering measured by Laval is less than one-half that given by formula (10), the Compton scattering being subtracted (cf. table 1). In the case of diamond, calcite and aluminium, the weak scattering, as measured by Laval, is even less, for small values of $(\sin\chi/2)/\lambda$, than the estimated Compton scattering. The reason for this is obscure, unless it can be in some way related to the existence of extinction effects for the comparatively perfect crystal faces used in the Laval scattering measurements.

By plotting the intensity of diffuse scattering power in reciprocal space, either theoretically, from the Jahn formula, say, or experimentally, as Laval and Lonsdale and Smith have done, it is at once apparent where the *minimum* scattering must occur. It is to be found near to the origin and also half way between lattice points such as the 002 and 004 for KCl, or between 002 and 024; (between 002 and 022 there is a tendency to a "bridge" of strong scattering, and this is even more marked between 004 and 024). It would appear that in one instance, at least, an unexpectedly low value of the scattering

(again about one-half of that given by formula (10)) was found in experiments on the Debye-Jauncey scattering. McNatt (1939), in measuring the diffuse scattering from Zn single crystals using monochromatic (Cu $K\alpha$ and Zn $K\alpha$) radiation, set the crystal so that the (0001) plane was at $\chi/2$ glancing angle to the incident beam, and the ionization chamber at various positions χ corresponding to $0{\cdot}14<(\sin\chi/2)/\lambda<0{\cdot}6$, avoiding all settings within 4° of a Bragg reflecting position. For $(\sin\chi/2)/\lambda=0{\cdot}24\rightarrow0{\cdot}37$ a dip or valley in the diffuse scattering curve was found, which attained its minimum value at about $(\sin\chi/2)/\lambda=0{\cdot}3$. He did not find a corresponding dip when the (0001) plane was horizontal

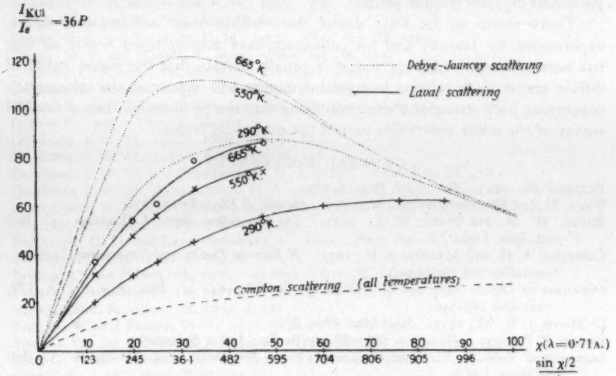


Figure 9. Comparison of Debye-Jauncey calculated and Laval observed weak diffuse scattering for KCl (Compton scattering included in each case).

(although his measurements appear to have fluctuated slightly from a smooth curve), and he also remarks that 'in previous experiments the use of general radiation would have smeared out fine structure effects so that they could not be observed. He considered whether the effect could be similar to the background discontinuities described earlier by Laval (1935), but these had been shown by Harvey (1939) to be due to harmonic wave-lengths in the monochromatic incident beam, while McNatt's effect persisted when the operating voltage was reduced below that required to excite even the half wave-length. He also considered other possible causes of the effect, such as a decrease of diffuse scattering when a strong Bragg reflection was taking place at some other angle, but tests showed that no such form of extinction occurs. McNatt came to the conclusion that the low value of the diffuse scattering which he observed was due to a genuine dip in the f-curve for $Zn_{v=0}$ in the neighbourhood of $(\sin\chi/2)/\lambda=0{\cdot}3$. and that it signified some deep-seated atomic distortion in the Zn electron cloud. This is not so. The position of the crystal in McNatt's experiments was such that he was, in fact, measuring the true "weak scattering"

(i.e. its minimum value) at a point where it could be observed, namely, half-way between the 0002 (i.e. $(\sin\chi/2)/\lambda = 0{\cdot}203$) and 0004 (i.e. $(\sin\chi/2)/\lambda = 0{\cdot}406$) reciprocal lattice points. Similar experiments carried out on any other crystals should give a similar dip in the diffuse scattering curve, provided that the crystal is suitably placed, monochromatic radiation used (a method involving the use of filters would not be satisfactory here) and narrow primary and scattered beams employed. McNatt (1940) found no dip in the atomic scattering curve for single crystal sections of Cu, because his sections were not parallel to any particular crystallographic plane.

There seems to be little doubt that, within their self-imposed limits, experiments by Jauncey and his colleagues have accomplished nearly all that has been claimed for them; but it is equally certain that the future study of diffuse scattering in relation to crystal dynamics will require all the information concerning both strong and weak scattering that can be obtained from a detailed survey of the *whole* observable part of the reciprocal lattice.

GENERAL REFERENCES

Blake, F. C., 1933. *Rev. Mod. Phys.* **5**, 169.

Born, M. and Göppert-Mayer, M., 1933. *Handb. d. Phys.* **24** (2), 638.

Bragg, W. H. and Bragg, W. L., 1933. *The Crystalline State*, **1** (London : G. Bell and Sons, Ltd.).

Compton, A. H. and Allison, S. K., 1935. *X Rays in Theory and Experiment* (London : Macmillan and Co., Ltd.).

Discussion on Diffuse Reflection of X Rays by Crystals, 1941–2. *Proc. Roy. Soc.* A, **179**, 1–101 and 289–340.

Du Mond, J. W. M., 1933. *Rev. Mod. Phys.* **5**, 1.

Ewald, P. P., 1933. *Handb. d. Phys.* **23** (2) (Berlin : Julius Springer).

International Tables for the Determination of Crystal Structures, 1935, **2** (London : G. Bell and Sons, Ltd.).

v. Laue, M., 1941. *Phys. u. Chem.*, **6**. "*Röntgenstrahl-Interferenzen*" (Leipzig : Akad. Verlags. Becker u. Erler Kom. Ges.).

Lonsdale, K., 1942. *Ann. Rep. Chem. Soc.* **38**, 92.

Reports on Progress in Physics, 1934. **1**, 234 ; 1935. **2**, 307 ; 1936. **3**, 344 ; 1937. **4**, 332 ; 1939. **6**, 345.

Schmid, E. and Boas, W., 1935. *Kristallplastizität* (Berlin : Julius Springer).

Symposium of Papers on the Quantum Theory of X-Ray Reflection, 1941. *Proc. Indian Acad. Sci., Bangalore*, **14 A**, 1–458.

OTHER REFERENCES

Baltzer, O. J., 1941. *Phys. Rev.* **60**, 460.

Baltzer, O. J. and Jauncey, G. E. M., 1939. *Phys. Rev.* **56**, 289.

Baltzer, O. J. and McNatt, E. M., 1939. *Phys. Rev.* **55**, 237.

Baxter, A. and Brentano, J. C. M., 1937. *Phil. Mag.* **24**, 473.

Bertsch, C. V., 1936. *Phys. Rev.* **49**, 128.

Bhagavantam, S. and Bhimasenachar, J., 1940. *Proc. Indian Acad. Sci., Bangalore*, **12 A**, 337.

Bhagavantam, S. and Bhimasenachar, J., 1941 *Proc. Indian Acad. Sci., Bangalore*, **13 A**, 266.

Blanc, M. and Weigle, J., 1937. *Helv. Phys. Acta*, **10**, 495.

Bloch, F., 1934. *Phys. Rev.* **46**, 674.

Born, M., 1942. *Nature, Lond.*, **149**, 402 ; *Proc. Roy. Soc.* A, **180**, 347 ; *Proc. Phys. Soc.* **54**, 362.

BORN, M. and LONSDALE, K., 1942. *Nature, Lond.*, **150**, 490.
BORN, M. and SARGINSON, K., 1941. *Proc. Roy. Soc.* A, **179**, 69.
BORRMANN, G., 1936. *Ann. Phys., Lpz.*, **27**, 669.
BRENTANO, J. C. M., HONEYBURNE, J. and BERRY, J. K., 1939. *Proc. Phys. Soc.* **51**, 668.
BRIDGMAN, P. W., 1925. *Proc. Am. Acad. Sci.* **60**, 305.
BRIDGMAN, P. W., 1929. *Proc. Am. Acad. Sci.* **64**, 19.
BRILL, R., HERMANN, C. and PETERS, C., 1942. *Ann. Phys., Lpz.*, **41**, 37.
BRINDLEY, G. W., 1936. *Phil. Mag.* **21**, 790 ; *Proc. Leeds Phil. Lit. Soc.* **3**, 200 ; *Nature, Lond.*, **137**, 315.
BRINDLEY, G. W. and RIDLEY, P., 1937. *Nature, Lond.*, **140**, 461.
BRINDLEY, G. W. and RIDLEY, P., 1938. *Proc. Phys. Soc.* **50**, 96 and 757.
BRINDLEY, G. W. and RIDLEY, P., 1939. *Proc. Phys. Soc.* **51**, 73.
BRINDLEY, G. W. and SPIERS, F. W., 1935. *Phil. Mag.* **20**, 865.
BRUCE, W. A., 1938. *Phys. Rev.* **53**, 802.
BRUCE, W. A. and MCNATT, E. M., 1937. *Phys. Rev.* **51**, 1065.
COLBY, M. Y. and HARRIS, S., 1934. *Phys. Rev.* **46**, 445.
COULSON, C. A. and DUNCANSON, W. E., 1942. *Proc. Camb. Phil. Soc.* **38**, 100.
DAWTON, R. H. V. M., 1937. *Proc. Phys. Soc.* **49**, 294.
DAYAL, B., 1941. *Proc. Indian Acad. Sci., Bangalore*, **14 A**, 421.
DEBYE, P., 1937. *Phys. Z.* **38**, 161.
DUMOND, J. W. M., 1929. *Phys. Rev.* **33**, 643.
DUMOND, J. W. M., 1932. *Phys. Rev.* **39**, 166.
DUMOND, J. W. M. and KIRKPATRICK, H. A., 1937. *Phys. Rev.* **52**, 419.
DUMOND, J. W. M., KIRKPATRICK, H. A. and ALDEN, L., 1932. *Phys. Rev.* **40**, 165.
DUNCANSON, W. E. and COULSON, C. A., 1941. *Proc. Camb. Phil. Soc.* **37**, 406.
DURAND, M. A., 1936. *Phys. Rev.* **50**, 449.
EHRHARDT, G. H. and LARK-HOROVITZ, K., 1940. *Phys. Rev.* **57**, 603.
FINBAK, C., 1939. *Phys. Z.* **40**, 26.
FINBAK, C. and HASSEL, O., 1936. *Z. phys. Chem.* **32**, 130 and 433.
FINBAK, C. and HASSEL, O., 1937. *Z. phys. Chem.* **35**, 25.
FÖRSTERLING, K., 1920. *Z. Phys.* **2**, 172.
FOX, G. W. and FRASER, W. A., 1935. *Phys. Rev.* **47**, 899.
FOX, G. W. and FREDERICK, J. R., 1938. *Phys. Rev.* **53**, 135.
FOX, G. W. and STEBBINS, D. W., 1939. *Phys. Rev.* **55**, 405.
GOENS, E., 1936. *Phys. Z.* **37**, 321.
GOETZ, A. and JACOBS, R. B., 1937. *Phys. Rev.* **51**, 151.
GREGG, R. Q. and GINGRICH, N. S., 1941. *Phys. Rev.* **59**, 619.
GRIFFITH, R. L. and HARRIS, P. M., 1941. *Phys. Rev.* **59**, 110.
GRÜNEISEN, E. and GOENS, E., 1924. *Z. Phys.* **26**, 235.
HALL, J. H., 1942. *Phys. Rev.* **61**, 158.
HARVEY, G. G., 1939. *Phys. Rev.* **56**, 247.
HARVEY, G. G., WILLIAMS, P. S. and JAUNCEY, G. E. M., 1934. *Phys. Rev.* **46**, 365.
HENDRICKS, S. B., POSNJAK, E. and KRACEK, F. C., 1932. *J. Amer. Chem. Soc.* **54**, 2766.
HICKS, B., 1937. *Phys. Rev.* **52**, 436.
HICKS, B., 1940. *Phys. Rev.* **57**, 665.
HÖNL, H., 1933. *Ann. Phys., Lpz.*, **18**, 625.
HOWARD, R. A., 1938. *Phys. Rev.* **53**, 324 and 966.
HUGHES, E. W., 1941. *J. Amer. Chem. Soc.* **63**, 1737.
IONA (Jr.), M., 1941. *Phys. Rev.* **60**, 822.
JACOBS, R. B. and GOETZ, A., 1935. *Phys. Rev.* **47**, 94 and 257.
JAHN, H. A., 1941. *Nature, Lond.*, **147**, 511.
JAHN, H. A., 1942a. *Proc. Roy. Soc.* A, **179**, 320.
JAHN, H. A., 1942b. *Proc. Roy. Soc.* A, **180**, 476.
JAHN, H. A. and LONSDALE, K., 1942. *Phys. Rev.* **61**, 375.
JAMES, H. M. and JOHNSON, V. A., 1939. *Phys. Rev.* **56**, 119.
JAMES, R. W. and BRINDLEY, G. W., 1928. *Proc. Roy. Soc.* A, **121**, 155.
JAUNCEY, G. E. M., 1939. *Phys. Rev.* **56**, 644.

JAUNCEY, G. E. M., 1941. *Phys. Rev.* **59**, 454 ; *Nature, Lond.*, **147**, 146.
JAUNCEY, G. E. M. and BALTZER, O. J., 1940. *Phys. Rev.* **58**, 1116.
JAUNCEY, G. E. M. and BALTZER, O. J., 1941. *Phys. Rev.* **59**, 699.
JAUNCEY, G. E. M., BALTZER, O. J. and MILLER, D. C., 1941. *Phys. Rev.* **59**, 908.
JAUNCEY, G. E. M. and BRUCE, W. A., 1936. *Phys. Rev.* **50**, 408 and 413.
JAUNCEY, G. E. M. and BRUCE, W. A., 1937 a. *Phys. Rev.* **51**, 1062.
JAUNCEY, G. E. M. and BRUCE, W. A., 1937 b. *Phys. Rev.* **51**, 1067.
JAUNCEY, G. E. M. and BRUCE, W. A., 1938. *Phys. Rev.* **54**, 163.
JAUNCEY, G. E. M. and CLAUS, W. D., 1934. *Phys. Rev.* **46**, 941.
JAUNCEY, G. E. M. and COMPTON, A. H., 1927. *Nature, Lond.*, **120**, 549.
JAUNCEY, G. E. M. and DEMING, J. H., 1935 a. *Phys. Rev.* **48**, 462.
JAUNCEY, G. E. M. and DEMING, J. H., 1935 b. *Phys. Rev.* **48**, 577.
JAUNCEY, G. E. M. and HARVEY, G. G., 1940. *Phys. Rev.* **58**, 179.
JAUNCEY, G. E. M. and JAQUES, A. T., 1936. *Phys. Rev.* **50**, 672.
JAUNCEY, G. E. M. and MCNATT, E. M., 1939. *Phys. Rev.* **55**, 498.
JAUNCEY, G. E. M. and MAY, H. L., 1924. *Phys. Rev.* **23**, 128.
JAUNCEY, G. E. M. and WILLIAMS, P. S., 1933. *Phys. Rev.* **44**, 794.
JOHNSON, V. A., 1940. *Phys. Rev.* **57**, 613.
KAKIUCHI, Y., 1938. *Phys. Rev.* **54**, 772.
KELLERMANN, E. W., 1940. *Philos. Trans.* A, **238**, 513.
KELLERMANN, E. W., 1941. *Proc. Roy. Soc.* A, **178**, 17.
KIRKPATRICK, H. A. and DUMOND, J. W. M., 1938. *Phys. Rev.* **54**, 802.
KIRKPATRICK, P., 1941. *Phys. Rev.* **59**, 452.
KIRKPATRICK, P. and ROSS, P. A., 1933. *Phys. Rev.* **43**, 596.
KLAUER, F., 1935. *Phys. Z.* **36**, 208.
KNAGGS, I. E. and LONSDALE, K., 1942. *J. Chem. Soc.* 417.
KNAGGS, I. E., LONSDALE, K., MÖLLER, A. and UBBELOHDE, A. R., 1940. *Nature, Lond.*, **145**, 820.
KOHLER, M., 1939. *Ann. Phys., Lpz.*, **36**, 227.
KOSSEL, W., 1937. *Ergeb. exakt. Naturw.* **16**, 295.
KOSSEL, W. and VOGES, H., 1935. *Ann. Phys., Lpz.*, **23**, 677.
LARK-HOROVITZ, K. and EHRHARDT, C. H., 1939. *Phys. Rev.* **55**, 605.
v. LAUE, M., 1935. *Ann. Phys., Lpz.*, **23**, 705.
v. LAUE, M., 1937. *Ann. Phys., Lpz.*, **28**, 528.
LAVAL, J., 1935. *C.R. Acad. Sci., Paris*, **201**, 889.
LAVAL, J., 1938. *C.R. Acad. Sci., Paris*, **207**, 169.
LAVAL, J., 1939 a. *C.R. Acad. Sci., Paris*, **208**, 1512.
LAVAL, J., 1939 b. *Bull. Soc. Franç. Min.* **62**, 137.
LIPSON, H. and RILEY, D. P., 1943. *Nature, Lond.*, **151**, 250, 502
LONSDALE, K., 1940. *Nature, Lond.*, **146**, 806.
LONSDALE, K., 1941 a. *Proc. Roy. Soc.* A, **177**, 272.
LONSDALE, K., 1941 b. *Nature, Lond.*, **147**, 481.
LONSDALE, K., 1942 a. *Proc. Roy. Soc.* A, **179**, 315.
LONSDALE, K., 1942 b. *Nature, Lond.*, **149**, 698.
LONSDALE, K., 1942 c. *Proc. Phys. Soc.* **54**, 314.
LONSDALE, K., KNAGGS, I. E. and SMITH, H., 1940. *Nature, Lond.*, **146**, 322.
LONSDALE, K., ROBERTSON, J. M. and WOODWARD, I., 1941. *Proc. Roy. Soc.* A, **178**, 43.
LONSDALE, K. and SMITH, H., 1941 a. *Nature, Lond.*, **148**, 112, 257 ; *Phys. Rev.* **60**, 617.
LONSDALE, K. and SMITH, H., 1941 b. *Proc. Roy. Soc.* A, **179**, 8.
LONSDALE, K. and SMITH, H., 1941 c. *Nature, Lond.*, **148**, 628.
LONSDALE, K. and SMITH, H., 1942 a. *Nature, Lond.*, **149**, 21.
LONSDALE, K. and SMITH, H., 1942 b. *Nature, Lond.*, **149**, 402.
MCNATT, E. M., 1939. *Phys. Rev.* **56**, 406.
MCNATT, E. M., 1940. *Phys. Rev.*, **57**, 621.
MCNATT, E. M. and JAUNCEY, G. E. M., 1939. *Phys. Rev.* **55**, 237.
MILLER, R. D., 1937. *Phys. Rev.* **51**, 959.
MILLER, R. D. and FOSTER (Jr.), E. S., 1936. *Phys. Rev.* **50**, 417.
NEMEJCOVÁ, A. and BROŽ, J., 1938. *Phys. Rev.* **54**, 379.
NILAKANTAN, P. and NAYAR, P. G. N., 1941. *Proc. Indian Acad. Sci., Bangalore*, **14 A**, 450.
NISHIKAWA, S., SAKISAKA, Y. and SUMATO, I., 1933. *Phys. Rev.* **43**, 363.

NISHIKAWA, S., SAKISAKA, Y. and SUMATO, I., 1934. *Sci. Pap. Inst. Phys. Chem. Res., Tokyo*, **25**, 20.
PISHAROTY, P. R., 1940. *Proc. Indian Acad. Sci., Bangalore*, **12 A**, 208.
PISHAROTY, P. R., 1941. *Proc. Indian Acad. Sci., Bangalore*, **14 A**, 56, 377 and 434.
PISHAROTY, P. R. and SUBRAHMANIAN, R. V., 1941. *Proc. Indian Acad. Sci., Bangalore*, **14 A**, 439.
PRESTON, G. D., 1939. *Proc. Roy. Soc.* A, **172**, 116.
PRESTON, G. D., 1941. *Proc. Roy. Soc.* A, **179**, 1.
PRESTON, G. D., 1942. *Nature, Lond.*, **149**, 373.
QUIMBY, S. L. and SIEGEL, S., 1938. *Phys. Rev.* **54**, 293.
RAMAN, C. V., 1941 a. *Proc. Indian Acad. Sci., Bangalore*, **13 A**, 1.
RAMAN, C. V., 1941 b. *Curr. Sci.* **10**, 473.
RAMAN, C. V., 1941 c. *Proc. Indian Acad. Sci., Bangalore*, **14 A**, 317 and 332.
RAMAN, C. V., 1942 a. *Proc. Roy. Soc.* A, **179**, 289 and 302 ; *Nature, Lond.*, **150**, 366.
RAMAN, C. V., 1942 b. *Curr. Sci.*, **11**, 65.
RAMAN, C. V. and NATH, N., 1940. *Proc. Indian Acad. Sci., Bangalore*, **12 A**, 427.
RAMAN, C. V. and NILAKANTAN, P., 1940 a. *Curr. Sci.* **9**, 165; *Nature, Lond.*, **145**, 667 and **146**, 686.
RAMAN, C. V. and NILAKANTAN, P., 1940 b. *Proc. Indian Acad. Sci., Bangalore*, **11 A**, 379, 389 and 398 ; **12 A**, 141 ; *Nature, Lond.*, **146**, 523.
RAMAN, C. V. and NILAKANTAN, P., 1941 a. *Curr. Sci.* **10**, 241.
RAMAN, C. V. and NILAKANTAN, P., 1941 b. *Proc. Indian Acad. Sci., Bangalore*, **14 A**, 356; *Nature, Lond.*, **147**, 118.
RAMAN, C. V. and NILAKANTAN, P., 1941 c. *Phys. Rev.* **60**, 63.
RENNINGER, M., 1937. *Naturwissenschaften*, **25**, 43 ; *Z. Phys.* **106**, 141.
ROBERTSON, R., FOX, J. J. and MARTIN, A. E., 1934. *Philos. Trans.* A, **232**, 463.
ROBERTSON, R., FOX, J. J. and MARTIN, A. E., 1936. *Proc. Roy. Soc.* A, **157**, 579.
ROSS, P. A. and KIRKPATRICK, P., 1934. *Phys. Rev.* **45**, 223, and **46**, 668.
SARGINSON, K., 1942. *Proc. Roy. Soc.* A, **180**, 305.
SCHARWÄCHTER, W., 1937. *Phys. Z.* **38**, 165.
SHANKLAND, R. S., 1937. *Phys. Rev.* **52**, 414.
SIEGEL, S., 1941. *Phys. Rev.* **59**, 371.
SIEGEL, S. and ZACHARIASEN, W. H., 1940. *Phys. Rev.* **57**, 795.
SIRKAR, S. C. and BISHUI, B. M., 1941. *Science and Culture*, **7**, 314.
SIRKAR, S. C. and GUPTA, J., 1940. *Proc. Nat. Inst. Sci. India*, **6**, 705.
SOMMERFELD, A., 1937. *Ann. Phys., Lpz.*, **29**, 715.
SUBRAHMANIAN, R. V., 1941. *Proc. Indian Acad. Sci., Bangalore*, **14 A**, 421.
TEAGUE, D. S., 1941. *Bull. Amer. Phys. Soc.* **16** (6), 7.
TEAGUE, D. S., 1942. *Phys. Rev.* **62**, 179.
VENKATESWARAN, C. S., 1941. *Proc. Indian Acad. Sci., Bangalore*, **14 A**, 387, 395, 402, 409, 415 and 426.
VOGES, H., 1936. *Ann. Phys., Lpz.*, **27**, 694.
WADLUND, A. P. R., 1938. *Phys. Rev.* **53**, 843.
WALLER, I. and HARTREE, D. R., 1929. *Proc. Roy. Soc.* A, **124**, 119.
WALLER, I. and JAMES, R. W., 1927. *Proc. Roy. Soc.* A, **117**, 214.
WEIGLE, J. and SMITH, C. S., 1942. *Phys. Rev.* **61**, 32.
WILSON, A. J. C., 1943. *Nature, Lond.*, **151**, 562.
WOLLAN, E. O., 1938. *Phys. Rev.* **53**, 203.
WOLLAN, E. O. and HARVEY, G. G., 1937. *Phys. Rev.* **51**, 1054.
WRIGHT, S. J., 1930. *Proc. Roy. Soc.* A, **126**, 613.
ZACHARIASEN, W. H., 1940. *Phys. Rev.* **57**, 597.
ZACHARIASEN, W. H., 1941 a. *Phys. Rev.* **59**, 207.
ZACHARIASEN, W. H., 1941 b. *Phys. Rev.* **59**, 766.
ZACHARIASEN, W. H., 1941 c. *Phys. Rev.* **59**, 860.
ZACHARIASEN, W. H., 1941 d. *Phys. Rev.* **59**, 909.
ZENER, C., 1935. *Phys. Rev.* **48**, 573.
ZENER, C., 1936. *Phys. Rev.* **49**, 122.
ZENER, C. and BILINSKY, S., 1936. *Phys. Rev.* **50**, 101.

THEORETICAL INVESTIGATIONS ON THE RELATION BETWEEN CRYSTAL DYNAMICS AND X-RAY SCATTERING

By MAX BORN, F.R.S.,
Department of Natural Philosophy, The University, Edinburgh

§1. HISTORICAL INTRODUCTION

THE theory of the influence of thermal motion on the scattering of X rays in crystals has recently been stimulated by the discovery of the "extra spots" on Laue photographs. The experimental side of this discovery is described in a separate *Report* by Mrs. Lonsdale, who, after a careful study of the various types of scattering of X rays by crystals and of the specific phenomena of the extra spots, comes to the conclusion that they are produced by the thermal motion. Alternative theories (Preston, 1938–41 b; Bragg, 1940–41 c; Raman, 1941 a–42; Raman and Nilakantan, 1940 a–41; Raman and Nagendra Nath, 1940) are shown to be unable to account for the observations and are also hardly reconcilable with well-established facts concerning the dynamical properties of crystal lattices (Born, 1941; Born, Lonsdale, Preston and Smith, 1942; Born and Blackman, 1942); short critical remarks will be made at appropriate places in this *Report*.

The effect of the thermal motion on an X-ray beam traversing the crystal has been compared with the effect of the agitated surface of the sea on the image of the setting sun. There is no sharp reflection, but a diffuse ribbon of light stretching towards the observer. This diffusion is obviously produced by the innumerable waves of various length and direction.

In the same way a beam of X rays passing through a crystal will be affected by the atomic waves which can be considered as the elements of the thermal motion. Each atomic wave modifies the direction, amplitude, frequency and phase of the reflected X ray. The change of frequency produced by the Doppler effect is inconspicuous, as the X-ray frequencies are extremely large compared to the frequencies of the thermal (infra-red) waves; but the change of phase is essential, since it produces incoherence of the scattered X rays. The problem is to calculate the intensity of scattering for any direction of observation.

The first mathematical theory was given by Debye (1913). Following Einstein's original theory of specific heat he considered the atoms as oscillating independently about their equilibrium positions, and applied classical statistics to these oscillators. The main result of his calculation was a decrease of the intensity, but no influence on the sharpness of the Laue spots or Bragg reflections, and a general diffuse background. v. Laue (1913) discussed Debye's theory with respect to the form of the spots and extended it to hemihedral structures. Darwin (1914) investigated deviations from the simple Laue theory produced by the features of the experimental arrangement (finite width of slits, hetero-

geneous x rays) and gave in this connection a simple derivation of the influence of temperature.

An important improvement was made independently by Debye (1914) and Schrödinger (1914a), by taking into account the mutual forces between the atoms of the lattice; both used the method of normal co-ordinates, introduced into the theory of crystal lattices by Born and v. Kármán (1912, 13).

Schrödinger (1914a) investigated in detail a linear model (chain of particles) and found that the Laue spots become broadened and diffuse with increasing temperature, while the total intensity remains unchanged. His work was attacked by Kern (1914), who maintained that the boundary conditions at the ends of the chain had an influence; but Schrödinger showed in a second paper (Schrödinger, 1914b) that this is not the case. Debye (1914) developed the general theory for a three-dimensional lattice and applied quantum statistics to the normal co-ordinates; at that time the question of the zero energy in Planck's formula for the oscillator energy was much disputed, and Debye discussed its influence on the scattering. The results with respect to the intensity distribution of the scattered radiation were the same as in his elementary theory, apart from the meaning of the constants. The discrepancy between Schrödinger's and Debye's results for the sharpness of the spots is due to the fact that Schrödinger deals with the simple case of a finite linear chain and is able to derive rigorous formulae for the total scattered intensity, while Debye, for the three-dimensional case, is compelled to use approximations in which the lowest term is interpreted as Laue scattering, the rest as background. For "ideal conditions", i.e. infinite lattice and perfectly plane monochromatic incident wave, the Laue scattering defined in this way consists of infinitely high and infinitely sharp peaks, each with a finite total intensity (like a Dirac δ function) which decreases with increasing temperature. Debye has determined the remainder and stated that it represents a diffuse background; but he has omitted to discuss its properties. Modern theory has shown that it represents an intensity distribution also with infinite, though not sharp, peaks at the reciprocal lattice points (interpreted by Schrödinger as broadening of the Laue spots). It is this background which produces the "extra spots".

In an interesting paper by Brillouin (1922) the importance of the background is stressed, and its connection with crystal dynamics is recognized. Faxén (1923) was the first to show that there ought to be maxima in the background, and also to give a formula for their position. In the same year Waller (1923) found that Debye's method of using normal co-ordinates was incorrect and that, in consequence, the temperature factor had to be replaced by its square.

Waller extended his investigations to the background scattering in his brilliant dissertation (Uppsala, 1925), which contains the complete formulae for the most general type of lattices; nothing essential has been added to these formulae by later theoretical research, apart from improvements in the presentation and practical application.

The first of these improvements is due to v. Laue (1926), who found that the agreement between the observed and calculated temperature dependence was not very satisfactory and tried to improve it by taking account of the change of frequency, namely, the Doppler effect, produced by the reflections of the

incident x ray on the (moving) elastic waves of the crystal. But, as mentioned before, this effect is negligible, and Laue's result is practically the same as that obtained by Waller.

The theory up to this time had been based on semi-classical considerations, combining the interference of waves with statistical averaging performed after the manner of Bohr's quantum theory. The first systematic use of quantum mechanics, applied to the normal co-ordinates as oscillators, was made by Waller (1928), and resulted in a complete confirmation of his previous formula.

A very thorough study, based on wave mechanics, was published by Ott (1935) with the purpose of estimating the influence of electronic excitation levels on the scattering. Waller's formula was again confirmed. This could have been foreseen from well-known considerations, due to Placzek (1934), which show that scattering of high-frequency radiation by an atomic system can be described in terms of one single constant, the polarizability, of the system. Ott's paper contains another improvement on Debye's theory. A central point of this is the calculation of the thermal average of the amplitude of an elastic wave, and this average can be reduced to the average of a harmonic function of the amplitude of a single oscillator. The formula for this thermal average was given by Debye (1914) without satisfactory proof. Schrödinger (1914a) and Waller (1925) proved it for high temperatures (classical statistics), and the latter also made a not quite successful attempt for low temperatures (quantum statistics). The first rigorous, though complicated, derivation of the formula is due to Ott (1935), while a simpler derivation has been given by Born and Sarginson (1941).

After the experimental discovery of the extra spots a great number of papers have appeared in which the theory has been redeveloped. Most of them contain no new features with respect to the theoretical foundations and the general formulae, but they discuss the real physical content of these formulae and compare them with experiment (Laval, 1938, 39; Mauguin and Laval, 1939; Zachariasen, 1940, 41; Jauncey, 1939, 40; Jauncey and Harvey, 1940; Jauncey and Baltzer, 1940; Born and Sarginson, 1941; v. Laue, 1941; Jahn, 1941, 42; Sarginson, 1942; Weigle, 1941–42b; Weigle and Smith, 1942; Born, 1942a).

In this *Report* these papers will not be presented in their historical order, but the theory will be systematically developed, using a method which is strictly based on the principles of quantum theory and which is mathematically rigorous. The merits or shortcomings of previous publications will be discussed at the proper place in the development; but the reader need not refer to these papers in order to follow the argument, except for details concerning comparison with experiments.

§ 2. GENERAL SCATTERING FORMULA*

The scattering of any atomic system depends on the distribution of the energy levels and on the matrix element, M^0_{rs}, of the electric moment M^0 between the states r, s of the undisturbed system. The energy of each excited state can be considered as the sum of electronic excitation and atomic vibration (for a molecule also rotation), the latter forming a fine structure around the electronic levels.

* A list of symbols used is given in Appendix III to this *Report*.

In general, electronic energies are too large to be excited in thermal equilibrium to any great extent. But nevertheless they influence the scattering in two ways:

(i) They are excited by the incident radiation. This produces true absorption and re-emission, which may proceed via intermediate states (fluorescence). If the excited state is in the continuous part of the energy spectrum the process is a photo-electric effect, and if the binding energy is small relative to the x-ray quantum this can also be described as a (modified) Compton effect. The scattering phenomena due to these quantum processes are described in Mrs. Lonsdale's *Report*.

(ii) The virtual existence of the electronic states determines the scattering, which can be described as consisting of double quantum transitions, namely, from a vibrational state belonging to the electronic ground level to any intermediate (electronic plus vibrational) state and back to a vibrational state of the electronic ground level.

Placzek (1934) has shown that the net result of these double transitions can be formulated in this way: the scattered vibration can be considered as produced by a dipole of moment $M^{(1)}=\alpha\mathcal{E}$, where $\mathcal{E}$ is the electric field of the incident wave and α a constant, the polarizability of the system, which depends on the state of vibration. Ott's (1935) investigation mentioned in § 1 has confirmed this result for lattices by a detailed calculation. The scattered intensity for the transition between two vibrational states, n, n', is proportional to $|M^{(1)}_{nn'}|^2$, where $M^{(1)}_{nn'}$ is the matrix element of $M^{(1)}$, and to the Boltzmann factor $\exp(-E_n/kT)$, where E_n is the (vibrational) energy of the lower state.

The induced moment $M^{(1)}$ for the whole system can, in first approximation, be considered as composed of the contributions of the single atoms. In each of these the external field, $\mathcal{E}$, acts with a different phase, depending on its position; and a second relative retardation is introduced by the different distances of the scattering atoms from the observer. Let $\mathbf{r}_k$ be the position vectors of the atoms in equilibrium, $\mathbf{u}_k$ be the displacement vectors due to thermal motion, and

$$\mathbf{K}=\frac{2\pi}{\lambda}\mathbf{s},\quad \mathbf{K}'=\frac{2\pi}{\lambda}\mathbf{s}',\quad |\mathbf{s}|=|\mathbf{s}'|=1, \qquad \ldots\ldots(2.1)$$

the wave vectors of the incident and scattered wave (change of wave-length λ neglected). Also let

$$\mathbf{Q}=\mathbf{K}'-\mathbf{K}=\frac{2\pi}{\lambda}(\mathbf{s}'-\mathbf{s}),\quad Q=|\mathbf{Q}|=\frac{4\pi}{\lambda}\sin\tfrac{1}{2}\chi, \qquad \ldots\ldots(2.2)$$

where χ is the angle of deflection; $\cos\chi=\mathbf{s}.\mathbf{s}'$. Then a harmonic wave $\mathcal{E}=\mathcal{E}_0\exp\{i\omega(t-\mathbf{r}.\mathbf{K}/c)\}$ produces the retarded moment

$$M^{(1)}=\alpha\mathcal{E}=\mathcal{E}_0e^{i\omega t}\sum_k\alpha_k(\mathbf{u})e^{i\mathbf{Q}.(\mathbf{r}_k+\mathbf{u}_k)}. \qquad \ldots\ldots(2.3)$$

The α_k represent the polarizabilities of the single atoms in the system. These may be not identical with those of the isolated atoms, but depend on the positions of the atoms, as indicated by writing $\alpha_k(\mathbf{u})$, in relation to neighbouring atoms and their modifying effects. The effect produced by this interaction is the analogue of the ordinary optical Raman effect (see § 13). At present the α_k will be considered as constants, and we write $\alpha_k=\alpha_0/f_k^0$, where α_0 is the polarizability of a free electron, f_k^0 the structure, or scattering, factor of th atom k (for $T=0$).

If the total scattering factor is defined by

$$f = \sum_k f_k^0 e^{i\mathfrak{Q}\cdot(\mathfrak{r}_k + \mathfrak{a}_k)}, \qquad \ldots\ldots(2.4)$$

and the incident intensity denoted by $I_0 = \mathfrak{S}_0^2$, then

$$|M_{nn'}^{(1)}|^2 = I_0 \alpha^2 |f_{nn'}|^2. \qquad \ldots\ldots(2.5)$$

To obtain the total scattered intensity, it is necessary to multiply by the Boltzmann factor of the initial state and by Hertz's well-known expression for the scattering due to a dipole of unit moment (which depends on the angle χ between the incident and scattered radiation), then to sum over all possible vibrational transitions. We then obtain

$$I = I_0 \sigma_0 \sum_n \sum_{n'} |f_{nn'}|^2 W_{n'}, \quad W_n = \frac{e^{-E_n/kT}}{\sum_{n'} e^{-E_{n'}/kT}}, \qquad \ldots\ldots(2.6)$$

where E_n is the vibrational energy in the state n and W_n the probability of finding the system in this state; further, with the usual notation, the Thomson factor

$$\sigma_0 = \left(\frac{e^2}{mc^2}\frac{1}{r}\right)^2 \tfrac{1}{2}(1 + \cos^2\chi), \qquad \ldots\ldots(2.7)$$

where χ is the same angle as in (2.2).

The double summation in (2.6) can be transformed into a single sum by virtue of the eigen functions $\psi_n(\xi)$ of an oscillator, or a system of oscillators, being real. This implies that

$$f_{nn'} = \int \psi_n(\xi) f(\xi) \psi_{n'}(\xi) d\xi = f_{n'n},$$

hence† $$\sum_{n'} |f_{nn'}|^2 = \sum_{n'} f_{nn'} f_{n'n}^* = |f|_{nn}^2,$$

and from (2.6) the scattering power reduces to

$$\sigma = \frac{I}{I_0} = \sigma_0 \sum_n |f|_{nn}^2 W_n. \qquad \ldots\ldots(2.8)$$

This formula reduces the scattering problem to the determination of the energy levels and the diagonal matrix elements of $|f|^2$, where f is defined by (2.4).

§ 3. SCATTERING BY A MOLECULE

The lattice is a coupled system of atoms with certain periodic properties. It is convenient, in the first place, to disregard these special properties and to derive the scattering formulae for any system of atoms.

The results suitably specialized can be applied either to molecules or to lattices.

Let the number of atoms be n, the masses $m_k (k = 1, 2, \ldots n)$. We introduce the abbreviation

$$A_k = f_k^0 e^{i\mathfrak{Q}\cdot\mathfrak{r}_k}, \qquad \ldots\ldots(3.1)$$

and can write (2.4)

$$f = \sum_k A_k e^{i\mathfrak{Q}\cdot\mathfrak{a}_k}. \qquad \ldots\ldots(3.2)$$

Then

$$|f|^2 = \sum_{kk'} A_k A_{k'}^* e^{i\mathfrak{Q}\cdot(\mathfrak{a}_k - \mathfrak{a}_{k'})}, \qquad \ldots\ldots(3.3)$$

† f^* means the conjugate complex of f.

In order to get rid of the mutual interactions we introduce normal co-ordinates $\xi_j (j = 1, 2, \ldots 3n)$, writing

$$\mathbf{u}_k m_k^{\frac{1}{2}} = \sum_j \mathbf{e}_k^j \xi_j. \qquad \ldots\ldots(3.4)$$

Here $\mathbf{e}_k^j / m_k^{\frac{1}{2}}$ is the amplitude vector of the atom k for the vibration j, frequency ω_j; they form an ortho-normal set in the $3n$-dimensional space, according to

$$\sum_k \mathbf{e}_k^j \cdot \mathbf{e}_k^{j'} = \delta_{jj'}. \qquad \ldots\ldots(3.5)$$

Now (3.3) becomes

$$|f|^2 = \sum_{kk'} A_k A_{k'}^* e^{i \sum_j \mu_{kk'}^j \xi_j}, \qquad \ldots\ldots(3.6)$$

where

$$\mu_{kk'}^j = \mathbf{Q} \cdot (\mathbf{e}_k^j / m_k^{\frac{1}{2}} - \mathbf{e}_{k'}^j / m_{k'}^{\frac{1}{2}}). \qquad \ldots\ldots(3.7)$$

The exponential in (3.6) can be written as a product

$$e^{i \sum_j \mu_{kk'}^j \xi_j} = \prod_j e^{i \mu_{kk'}^j \xi_j} \qquad \ldots\ldots(3.8)$$

and as the vibrational energy of a system of oscillators is a sum

$$E = \sum_n \epsilon_{n_j}, \quad \epsilon_{n_j} = \hbar \omega_j (n_j + \tfrac{1}{2}), \qquad \ldots\ldots(3.9)$$

the probability (2.6) is also a product

$$W_n = \prod_j w_{n_j}, \quad w_{n_j} = \frac{e^{-\epsilon_{n_j}/kT}}{\sum_{n_j} e^{-\epsilon_{n_j}/kT}}. \qquad \ldots\ldots(3.10)$$

Introducing (3.6) into (2.8) and using (3.8) and (3.10) one has

$$\sigma = \sigma_0 \sum_{kk'} A_k A_{k'}^* \prod_j \sum_{n_j} (e^{i \mu_{kk'}^j \xi_j})_{n_j n_j} \cdot w_{n_j}. \qquad \ldots\ldots(3.11)$$

Thus the scattering problem is reduced to the determination of the thermal average

$$(e^{i\mu\xi})_{\mathrm{av}} = \sum_n (e^{i\mu\xi})_{nn} w_n = \frac{\sum_n (e^{i\mu\xi})_{nn} e^{-\epsilon_n/kT}}{\sum_n e^{-\epsilon_n/kT}}, \qquad \ldots\ldots(3.12)$$

where ξ is the amplitude and ϵ the energy of a linear oscillator. Debye (1913, 14; see § 1) replaced $\exp(i\mu\xi)$ by the first terms of the expansion $1 + i\mu\xi - \frac{1}{2}\mu^2\xi^2 + \ldots$, and averaging term by term obtained $(\xi)_{\mathrm{av}} = 0$, $(\xi^2)_{\mathrm{av}} = \epsilon/\omega^2$, where ω is the frequency and ϵ the energy of the oscillator. Hence

$$\{\exp(i\mu\xi)\}_{\mathrm{av}} = 1 - \tfrac{1}{2}\mu^2\bar{\epsilon}/\omega^2 + \ldots,$$

and this is approximately equal to $\exp\{-\frac{1}{2}\mu^2\bar{\epsilon}/\omega^2\}$. This is, of course, no proof; as explained in § 1, a rigorous derivation of the formula

$$(e^{i\mu\xi})_{\mathrm{av}} = e^{-\frac{1}{2}\mu^2\bar{\epsilon}/\omega^2} \qquad \ldots\ldots(3.13)$$

was given by Ott (1935) and a simpler one by Born and Sarginson (1941) which is reproduced in Appendix I to this *Report*. Using this result (3.11) becomes

$$\sigma = \sigma_0 \sum_{kk'} A_k A_{k'}^* \prod_j \exp\{-\tfrac{1}{2}(\mu_{kk'}^j)^2 \bar{\epsilon}_j / \omega_j^2\}. \qquad \ldots\ldots(3.14)$$

or, if the notation

$$\tfrac{1}{2} \sum (\mu_{kk'}^j)^2 \bar{\epsilon}_j / \omega_j^2 = U_{kk'} \qquad \ldots\ldots(3.15)$$

is introduced,

$$\sigma = \sigma_0 \sum_{kk'} A_k A_{k'}^* e^{-U_{kk'}}. \qquad \ldots\ldots(3.16)$$

Before proceeding, it is necessary to draw attention to the fact that in systems such as crystal lattices a great number of frequencies are small and approach zero. For these the exponent in (3.14) is large, and although the formula represents the effect of the small thermal oscillations, it is not possible to simplify it by expanding the exponential function. However, many of the modern treatments of theory quoted in § 1 use this expansion from the beginning and can not therefore be considered as satisfactory. The only correct way is to use Ott's formula (3.13) and to transform (3.11) into (3.16); this expression can be expanded in powers of $U_{kk'}$, for this sum defined by (3.15) is convergent in the case of a lattice in spite of the existence of frequency branches approaching zero, and its value is small in all practical cases.

We have now to investigate $U_{kk'}$; substituting $\mu_{kk'}^j$ from (3.7) one has

$$U_{kk'} = \tfrac{1}{2} \sum_j (\tilde{\epsilon}_j/\omega_j^2)[\mathbf{Q} \cdot (\mathbf{e}_k^j/m_k^{\frac{1}{2}} - \mathbf{e}_{k'}^j/m_{k'}^{\frac{1}{2}})]^2; \qquad \ldots\ldots(3.17)$$

writing the scalar product as a sum, $\mathbf{a}.\mathbf{b} = \sum_x a_x b_x$, and changing the order of summation, one gets

$$U_{kk'} = \tfrac{1}{2} \sum_{xy} Q_x Q_y \sum_j \frac{\tilde{\epsilon}_j}{\omega_j^2} \left\{ \frac{e_{kx}^j e_{ky}^j}{m_k} + \frac{e_{k'x}^j e_{k'y}^j}{m_{k'}} - \frac{e_{kx}^j e_{k'y}^j + e_{ky}^j e_{k'x}^j}{(m_k m_{k'})^{\frac{1}{2}}} \right\}. \qquad \ldots\ldots(3.18)$$

The sums appearing here can be simplified by using a well-known algebraic theorem, applied to eigenvectors of the mechanical systems, in the manner to be described in the following section.

§ 4. THE DYNAMICAL AND THE SCATTERING MATRIX OF A MOLECULE

It is convenient to use the reduced amplitudes $\mathbf{v}_k = \mathbf{u}_k m_k^{\frac{1}{2}}$ already introduced in the definition (3.4) of the eigenvectors $\mathbf{e}_k^j$. The potential energy Φ (eigenvalue of the electronic energy for stationary nuclei in the lowest state) can be developed in powers of the $\mathbf{v}_k$; the linear terms vanish in equilibrium, and the second order terms have the coefficients

$$\frac{1}{(m_k m_{k'})^{\frac{1}{2}}} \left(\frac{\partial^2 \Phi}{\partial x_k \partial y_{k'}} \right)_0 = \mathscr{D}_{\substack{kk' \\ xy}}, \qquad \ldots\ldots(4.1)$$

where the suffix 0 denotes the equilibrium values. These quantities are the elements of the *dynamical matrix* $\mathscr{D}$ in the $3n$-dimensional space with the co-ordinates

$$1x\ 1y\ 1z\ 2x\ 2y\ 2z\ 3x \ldots\ldots nx\ ny\ nz. \qquad \ldots\ldots(4.2)$$

The kinetic and potential energy are

$$T = \tfrac{1}{2} \sum_k \dot{\mathbf{v}}_k^2 = \tfrac{1}{2} \tilde{\dot{v}} \dot{v}, \qquad \Phi = \tfrac{1}{2} \sum_{kx} \sum_{k'y} \mathscr{D}_{\substack{kk' \\ xy}} v_{kx} v_{k'y} = \tfrac{1}{2} \tilde{v} \mathscr{D} v; \qquad \ldots\ldots(4.3)$$

in the second form given v represents a one-column matrix, and $\tilde{v}$ the transposed —i.e. a one-row matrix.

The equations of motion for small vibrations are

$$\ddot{v}_{kx} + \sum_{k'y} \mathscr{D}_{\substack{kk' \\ xy}} v_{k'y} = 0, \quad \text{or} \quad \ddot{v} + \mathscr{D} v = 0. \qquad \ldots\ldots(4.4)$$

If the whole system is moved as a rigid body, i.e. if all $\mathbf{u}_k = \mathbf{v}_k/m_k^{\frac{1}{2}}$ are equal, the forces must vanish; hence

$$\sum_{k'} m_{k'}^{\frac{1}{2}} \mathscr{D}_{\substack{kk'\\ xy}} = 0. \qquad \text{.....(4.5)}$$

For a vibration of frequency ω_j we write $\mathbf{v}_k = \mathbf{e}_k^j$, and have

$$\omega_j^2 e_{kx}^j = \sum_{k'y} \mathscr{D}_{\substack{kk'\\ xy}} e_{k'y}^j, \quad (j = 1, 2, \ldots . 3n). \qquad \text{.....(4.6)}$$

One can now form the square of the matrix $\mathscr{D}$, $\mathscr{D}^2$, with the elements

$$(\mathscr{D}^2)_{\substack{kk'\\ xy}} = \sum_{l z} \mathscr{D}_{\substack{kl\\ xz}} \mathscr{D}_{\substack{lk'\\ zy}},$$

and similarly for higher powers of $\mathscr{D}$. Defining $\mathscr{D}^{-1}$ by the matrix equation $\mathscr{D} . \mathscr{D}^{-1} = 1$, where 1 is the $3n$-dimensional unit matrix, one can also define $\mathscr{D}^p$ for any negative p. Hence a function $f(\mathscr{D})$ can be defined as a matrix represented by a power series (with positive and/or negative exponents) of $\mathscr{D}$. Then the following well-known theorem holds (a simple proof has been given by Born, 1942a; see Appendix II to this *Report*):

$$\sum_j f(\omega_j^2) e_{kx}^j e_{k'y}^j = [f(\mathscr{D})]_{\substack{kk'\\ xy}}, \qquad \text{.....(4.7)}$$

or

$$\sum_j \mathbf{e}^j f \mathbf{e}^j = f(\mathscr{D}).$$

(3.18) consists of four sums of this form for the special function

$$\phi(\omega^2) = \frac{\bar{\epsilon}}{\omega^2} = \frac{\hbar}{\omega}\left(\frac{1}{2} + \frac{1}{e^{\hbar\omega/kT} - 1}\right) = \frac{\hbar^2}{kT} \frac{\coth \frac{1}{2}\beta}{2\beta}, \quad \beta = \frac{\hbar\omega}{kT}. \qquad \text{.....(4.8)}$$

This can, indeed, be expanded in a power series in ω, namely,

$$\left.\begin{aligned} \phi(\omega^2) &= \frac{\hbar^2}{kT}\left\{\beta^{-2} - \sum_{n=1}^{\infty} (-1)^n \frac{B_n}{(2n)!} \beta^{2(n-1)}\right\} \\ &= kT\left\{\frac{1}{\omega^2} - \sum_{n=1}^{\infty} (-1)^n \frac{B_n}{(2n)!} \left(\frac{\hbar}{kT}\right)^{2n} \omega^{2(n-1)}\right\} \end{aligned}\right\}, \qquad \text{.....(4.9)}$$

where B_n are the Bernoulli numbers: $B_1 = \frac{1}{6}$, $B_2 = \frac{1}{30}$, $B_3 = \frac{1}{42}$, ...

The matrix obtained by forming this function of $\mathscr{D}$,

$$\mathscr{S} = \phi(\mathscr{D}) = \frac{\hbar}{2} \mathscr{D}^{-1} \coth\left(\frac{\hbar \mathscr{D}^{\frac{1}{2}}}{2kT}\right) = kT\left\{\mathscr{D}^{-1} + \frac{1}{12}\left(\frac{\hbar}{kT}\right)^2 - \frac{1}{720}\left(\frac{\hbar}{kT}\right)^4 \mathscr{D} + \ldots .\right\}. \qquad \text{.....(4.10)}$$

will be called the *scattering matrix*.

§ 5. SINGLE AND DOUBLE SCATTERING OF A MOLECULE

Using (4.7) and (4.10) one can write (3.18) as sum of 3 terms

$$U_{kk'} = M_k + M_{k'} - M_{kk'}, \qquad \text{.....(5.1)}$$

where

$$\left.\begin{aligned} M_{kk'} &= \frac{1}{(m_k m_{k'})^{\frac{1}{2}}} \sum_{xy} \mathscr{S}_{\substack{kk'\\ xy}} Q_x Q_y, \\ M_k &= \tfrac{1}{2} M_{kk}. \end{aligned}\right\} \qquad \text{.....(5.2)}$$

Substituting (5.1) into (3.16) one has

$$\sigma = \sigma_0 \sum_{kk'} A_k A_{k'}^* e^{-M_k - M_{k'} + M_{kk'}}, \qquad \text{.....(5.3)}$$

Although the three terms in the exponent are obviously of the same order of magnitude, it is convenient to expand only the last exponential factor into powers of the argument, writing $\exp M_{kk'} = 1 + M_{kk'} + \ldots\ldots$ (Weigle and Bleuler (1942) have recently shown how the theory can be developed without this approximation, and have given explicit formulae for the terms quadratic in $M_{kk'}$.) The limits of the validity of the expansion will be discussed later (§ 10). Keeping only the first-order term, one can split σ into two parts,

$$\sigma = \sigma_1 + \sigma_2. \qquad \ldots\ldots(5.4)$$

σ_1 represents the scattering by single atoms, modified, however, by the thermal agitation, *single scattering* for short, namely,

$$\sigma_1 = \sigma_0 \left|\sum_k A_k e^{-M_k}\right|^2 = \sigma_0 \left|\sum_k f_k^0 e^{i\mathbf{Q}\cdot\mathbf{r}_k - M_k}\right|^2 = \sigma_0 \left|\sum_k f_k e^{i\mathbf{Q}\cdot\mathbf{r}_k}\right|^2, \qquad \ldots\ldots(5.5)$$

where the definition (3.1) of A_k is used and

$$f_k = f_k^0 e^{-M_k} \qquad \ldots\ldots(5.6)$$

is a *modified scattering factor*, depending on temperature (as ϕ and, therefore, M_k depend on T).

The second term is the *double scattering* depending on all pairs of atoms:

$$\sigma_2 = \sigma_0 \sum_{kk'} A_k A_{k'}^* e^{-M_k - M_{k'}} M_{kk'} = \sigma_0 \sum_{kk'} f_k f_{k'} M_{kk'} e^{i\mathbf{Q}\cdot(\mathbf{r}_k - \mathbf{r}_{k'})}. \qquad \ldots\ldots(5.7)$$

Debye (1927) was the first to direct attention to the possibility of investigating the structure of molecules by the scattering of x rays. He used the formula (5.5) without the temperature factor (f_k^0 instead of f_k), averaged it over all orientations of the molecule and showed that an observable effect (interference rings) could be expected; this effect he actually found by experiment. The present theory generalizes Debye's calculations by taking the thermal motion into account. For working out these effects in detail one should use the rigorous formula (5.3) rather than the approximations (5.4), (5.5), (5.7); but such detailed treatment is not the purpose of this *Report*.

Substituting (5.2) in (5.7) an expression of the form†

$$\sigma_2 = \sigma_0 \sum_{kx} \sum_{k'y} \mathscr{S}_{\substack{kk' \\ xy}} w_{kx} w_{k'y}^* = \sigma_0 (w^+ \mathscr{S} w) \qquad \ldots\ldots(5.8)$$

is obtained, where $w_{kx}, \ldots$ are the components of the vector

$$\mathbf{w}_k = \frac{f_k}{m_k^{\frac{1}{2}}} e^{i\mathbf{Q}\cdot\mathbf{r}_k}\mathbf{Q}. \qquad \ldots\ldots(5.9)$$

In the second expression (5.8), $w_{1x}, w_{1y}, \ldots w_{nz}$ are taken as a one-column matrix w, and w^+ means the conjugate transposed form—i.e. a one-row matrix. (5.8) justifies the term "scattering matrix" for $\mathscr{S} = \phi(\mathscr{D})$; it determines the (double) scattering intensity as a Hermitean form of the $\mathbf{w}_k$ in a way similar to that in which the dynamical matrix $\mathscr{D}$ determines the potential energy as a function of the (reduced) displacements $\mathbf{v}_k$.

§ 6. SCATTERING BY A LATTICE

The particular application of the scattering formulae (5.5), (5.8) to lattices is now extremely simple.

† σ_2 can also be expressed as quadric in the components of $\mathbf{Q}$; but this is of no use, as the coefficients depend again on the same variables $\mathbf{Q}$.

Let the basis of the lattice be given by the vectors $\mathbf{a}_1$, $\mathbf{a}_2$, $\mathbf{a}_3$ of the cell, and the position vectors $\mathbf{r}_k$ of the n points in the cell, $k=1, 2, \ldots n$. An arbitrary lattice point is then given by (regarding notation, see Born, 1923)

$$\mathbf{r}_k^l = \mathbf{r}_k + \mathbf{r}^l, \quad \mathbf{r}^l = \mathbf{a}_1 l_1 + \mathbf{a}_2 l_2 + \mathbf{a}_3 l_3, \qquad \ldots\ldots(6.1)$$

where the cell index l represents three integers l_1, l_2, l_3. Wherever the index k appears in the preceding formulae it has to be replaced by (k, l). The space of the dynamical variables $\mathbf{v}_k^l$ has $3nN$ dimensions, where N is the number of cells in the crystal.

The elements of the dynamical matrix $\mathscr{D}$ depend only on the difference of the indices $l-l'$, by virtue of the periodicity of the lattice; introducing a quantity $\mathscr{D}_{\substack{l \\ kk' \\ xy}}$ the elements can be written $\mathscr{D}_{\substack{l-l' \\ kk' \\ xy}}$. The equations (4.5) are now

$$\sum_{k'} m_k^{\frac{1}{2}} \sum_{l} \mathscr{D}_{\substack{l \\ kk' \\ xy}} = 0. \qquad \ldots\ldots(6.2)$$

It is convenient to consider $\mathscr{D}_{\substack{l \\ kk' \\ xy}}$ as elements of a $3n$-dimensional matrix $\mathscr{D}_l$ in the sub-space with co-ordinates $1x$, $1y$, $1z$, $2x$, $\ldots\ldots$ ny, nz.

In the case of a one-dimensional lattice the matrix $\mathscr{D}$ can be written as a "circulant".

$$\mathscr{D} = \begin{pmatrix} \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \mathscr{D}_{-2} & \mathscr{D}_{-1} & \mathscr{D}_0 & \mathscr{D}_1 & \mathscr{D}_2 & \mathscr{D}_3 & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \mathscr{D}_{-2} & \mathscr{D}_{-1} & \mathscr{D}_0 & \mathscr{D}_1 & \mathscr{D}_2 & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \mathscr{D}_{-3} & \mathscr{D}_{-2} & \mathscr{D}_{-1} & \mathscr{D}_0 & \mathscr{D}_1 & \mathscr{D}_2 & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \end{pmatrix} \qquad \ldots\ldots(6.3\,a)$$

For a three-dimensional lattice where l represents three indices such a simple representation is actually not possible, but can be used symbolically.

The simplest example is a monatomic one-dimensional lattice (chain of equidistant equal particles); here the $\mathscr{D}_l$ are not matrices but numbers, and $m\mathscr{D}_l(u_{p+l}-u_p)$ represents the quasi-elastic force between any particle (p) and that particle which is l units away from it. If only first neighbours are acting on one another with a force a per unit change of distance, one has

$$\mathscr{D} = \begin{pmatrix} \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & 0 & 0 & \mathscr{D}_{-1} & \mathscr{D}_0 & \mathscr{D}_1 & 0 & 0 & \cdot & \cdot & \cdot \\ \cdot & \cdot & 0 & 0 & 0 & \mathscr{D}_{-1} & \mathscr{D}_0 & \mathscr{D}_1 & 0 & 0 & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \end{pmatrix} \qquad \begin{aligned} \mathscr{D}_1 = \mathscr{D}_{-1} &= -\frac{a}{m}, \\ \mathscr{D}_0 &= \frac{2a}{m}. \end{aligned} \qquad \ldots\ldots(6.3\,b)$$

For the diatomic linear chain the $\mathscr{D}_l$ are matrices of two rows and two columns. If the chain is equidistant with only first neighbours acting on one another (force constant a) $\mathscr{D}$ has the same form (6.3 *b*), where

$$\mathscr{D}_{-1} = \begin{pmatrix} 0 & \frac{a}{(m_1 m_2)^{\frac{1}{2}}} \\ 0 & 0 \end{pmatrix}, \quad \mathscr{D}_0 = \begin{pmatrix} \frac{2a}{m_1} & \frac{-a}{(m_1 m_2)^{\frac{1}{2}}} \\ \frac{-a}{(m_1 m_2)^{\frac{1}{2}}} & \frac{2a}{m_2} \end{pmatrix}, \quad \mathscr{D}_1 = \begin{pmatrix} 0 & 0 \\ \frac{-a}{(m_1 m_2)^{\frac{1}{2}}} & 0 \end{pmatrix} \qquad \ldots\ldots(6.3\,c)$$

This follows from the fact that the equations of motion

$$m_1\ddot{u}_1^l = \alpha(u_2^{l-1} - u_1^l) - \alpha(u_1^l - u_2^l),$$
$$m_2\ddot{u}_2^l = \alpha(u_1^l - u_2^l) - \alpha(u_2^l - u_2^{l+1})$$

can be written as equations for the variables $v_k^l = u_k^l m_k^l$ in the matrix form

$$\ddot{v} + \mathscr{D}_{-1}v^{l-1} + \mathscr{D}_0 v^l + \mathscr{D}_1 v^{l+1} = 0,$$

where v^l means the column matrix $\begin{pmatrix} v_1^l \\ v_2^l \end{pmatrix}$.

For a simple three-dimensional lattice one has

$$\mathscr{D}_l = \begin{pmatrix} \mathscr{D}_{xx}^l & \mathscr{D}_{xy}^l & \mathscr{D}_{xz}^l \\ \mathscr{D}_{yx}^l & \mathscr{D}_{yy}^l & \mathscr{D}_{yz}^l \\ \mathscr{D}_{zx}^l & \mathscr{D}_{zy}^l & \mathscr{D}_{zz}^l \end{pmatrix} \qquad \ldots\ldots (6.3\,d)$$

Models of this type, namely simple cubic lattices, have been used by Born and v. Kármán (1912, 13) and later by Blackman (1935 a–42) for the calculation of specific heats and by Zachariasen (1940) for the thermal scattering of x rays.

Here the formulae for the *face-centred cubic lattice* will be given, as this has several advantages; e.g., if only forces between next neighbours are considered, the number of quasi-elastic constants is three, equal to that of the molar elastic constants, and the structure is stable even for central forces (where the number of constants reduces to one), while the simple lattice is unstable in this case (Born, 1940; Misra, 1940; Peng and Power, 1942). The point $l=0$, i.e. $l_1=l_2=l_3=0$, has twelve neighbours, namely :

l	1	2	3	4	5	6	7	8	9	10	11	12
l_1	0	0	1	−1	1	1	0	0	−1	1	−1	−1
l_2	1	1	0	0	1	−1	−1	−1	0	0	−1	1
l_3	1	−1	1	1	0	0	−1	1	−1	−1	0	0

By simple symmetry considerations the corresponding thirteen matrices are found to be

$$\mathscr{D}_0 = \frac{4}{m}(\alpha+2\beta)\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \qquad \left.\begin{matrix} \mathscr{D}_1 = \mathscr{D}_7 \\ \mathscr{D}_2 = \mathscr{D}_8 \end{matrix}\right\} = \frac{-1}{m}\begin{pmatrix} \alpha & 0 & 0 \\ 0 & \beta & \pm\gamma \\ 0 & \pm\gamma & \beta \end{pmatrix},$$

$$\left.\begin{matrix} \mathscr{D}_3 = \mathscr{D}_9 \\ \mathscr{D}_4 = \mathscr{D}_{10} \end{matrix}\right\} = \frac{-1}{m}\begin{pmatrix} \beta & 0 & \mp\gamma \\ 0 & \alpha & 0 \\ \mp\gamma & 0 & \beta \end{pmatrix} \qquad \left.\begin{matrix} \mathscr{D}_5 = \mathscr{D}_{11} \\ \mathscr{D}_6 = \mathscr{D}_{12} \end{matrix}\right\} = \frac{-1}{m}\begin{pmatrix} \beta & \pm\gamma & 0 \\ \pm\gamma & \beta & 0 \\ 0 & 0 & \alpha \end{pmatrix}$$

$\ldots\ldots (6.3\,e)$

The equations of motion are :

$$\ddot{\mathbf{v}}^0 + \mathscr{D}_0\mathbf{v}^0 + \mathscr{D}_1(\mathbf{v}^1 + \mathbf{v}^7) + \mathscr{D}_2(\mathbf{v}^2 + \mathbf{v}^8) + \ldots + \mathscr{D}_6(\mathbf{v}^6 + \mathbf{v}^{12}) = 0,$$

where $\mathbf{v}^0$ is a one-column matrix with the elements $v_x^{l_1 l_2 l_3}$, $v_y^{l_1 l_2 l_3}$, $v_z^{l_1 l_2 l_3}$ and $\mathbf{v}^1$ is a similar matrix with the elements $v_x^{l_1, l_2+1, l_3+1}, \ldots$, etc.

For a diatomic three-dimensional lattice $\mathscr{D}_l$ has six rows and columns :

$$\mathscr{D} = \begin{pmatrix} \mathscr{D}_{xx}^{11} & \mathscr{D}_{xy}^{11} & \mathscr{D}_{xz}^{11} & \mathscr{D}_{xx}^{12} & \mathscr{D}_{xy}^{12} & \mathscr{D}_{xz}^{12} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \mathscr{D}_{xx}^{21} & \mathscr{D}_{xy}^{21} & \mathscr{D}_{xz}^{21} & \mathscr{D}_{xx}^{22} & \mathscr{D}_{xy}^{22} & \mathscr{D}_{xz}^{22} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \end{pmatrix} \qquad \ldots\ldots (6.3\,f)$$

The only instance of detailed calculations for this case is that afforded by the work of Kellermann (1940) on NaCl.

There are certain quantities independent of l, e.g. the masses m_k and the scattering factors f_k^0, which are characteristic of the type of atom. The same holds for M_k defined by (5.2), and hence for f_k (5.6); for the $M_{kk'}$ are now to be written $M_{\substack{l-l' \\ kk'}}$; hence $M_k = \frac{1}{2}M_{\substack{0 \\ kk}}$ is now $\frac{1}{2}M_{\substack{0 \\ kk}}$, for which we can again write

$$M_k = \tfrac{1}{2}M_{\substack{0 \\ kk}} = \frac{1}{2m_k}\sum_{xy}\mathscr{D}_{\substack{0 \\ kk \\ xy}} Q_x Q_y. \qquad \ldots\ldots(6.4)$$

The single scattering (5.5) now represents the *Laue scattering*; it becomes

$$\sigma_1 = \sigma_0\left|\sum_{kl} f_k e^{i\mathbf{Q}.(\mathbf{r}_k+\mathbf{r}^l)}\right|^2 = \sigma_0\left|\sum_k f_k e^{i\mathbf{Q}.\mathbf{r}_k}\right|^2 . \left|\sum_l e^{i\mathbf{Q}.\mathbf{r}^l}\right|^2. \qquad \ldots\ldots(6.5)$$

If Q_1, Q_2, Q_3 are defined by

$$Q_\alpha = \mathbf{Q}.\mathbf{a}_\alpha, \quad \alpha = 1, 2, 3; \quad \mathbf{Q} = Q_1\mathbf{b}_1 + Q_2\mathbf{b}_2 + Q_3\mathbf{b}_3 = (Q\mathbf{b}), \qquad \ldots\ldots(6.6)$$

where $\mathbf{b}_1$, $\mathbf{b}_2$, $\mathbf{b}_3$ are the reciprocal lattice vectors, then

$$\mathbf{Q}.\mathbf{r}^l = Q_1 l_1 + Q_2 l_2 + Q_3 l_3 = (Ql). \qquad \ldots\ldots(6.7)$$

The last factor in (6.5) is that characteristic of Laue's theory and can be evaluated for a parallelepiped, similar to the cell:

$$\left|\sum_l e^{i\mathbf{Q}.\mathbf{r}^l}\right|^2 = \left|\sum_{l_1 l_2 l_3} e^{i(Q_1 l_1 + Q_2 l_2 + Q_3 l_3)}\right|^2 = \prod_{\alpha=1}^{3}\left|\sum_{l_\alpha=1}^{N_\alpha} e^{iQ_\alpha l_\alpha}\right|^2$$

$$= (2\pi)^3 N \prod_{\alpha=1}^{3}\left(\frac{1}{(2\pi N_\alpha)^{\frac{1}{2}}}\frac{\sin\frac{1}{2}N_\alpha Q_\alpha}{\sin\frac{1}{2}Q_\alpha}\right)^2 = (2\pi)^3 N\delta_N(Q), \qquad \ldots\ldots(6.8)$$

where $N_1 N_2 N_3 = N$. The function $\delta_N(Q) = \delta_{N_1}(Q_1)\delta_{N_2}(Q_2)\delta_{N_3}(Q_3)$ approaches, for N_1, N_2, $N_3 \to \infty$, a periodic Dirac function $\delta(Q)$ with period 2π for each Q_α, and $\int\delta(Q)d\overline{Q} = 1$, taken over a period. N_1, N_2, N_3 are assumed to be very large, so that $\delta(Q)$ can be written for $\delta_N(Q)$. Then

$$\sigma_1 = \sigma_0 N(2\pi)^3\delta(Q) \,.\, |F(Q)|^2, \qquad \ldots\ldots(6.9)$$

where

$$F(Q) = \sum_k f_k e^{i\mathbf{Q}.\mathbf{r}_k} = \sum_k f_k^0 e^{i\mathbf{Q}.\mathbf{r}_k - M_k} \qquad \ldots\ldots(6.10)$$

is the lattice structure factor modified by temperature (contained in M_k (see (6.4)).

The intensity distribution σ_1 has infinitely high and sharp peaks for

$$Q_\alpha = 2\pi h_\alpha, \quad \text{or} \quad \mathbf{Q} = \mathbf{K}' - \mathbf{K} = \mathbf{Q}^{(h)}, \qquad \ldots\ldots(6.11)$$

where $\mathbf{Q}^{(h)} = 2\pi(h_1\mathbf{b}_1 + h_2\mathbf{b}_2 + h_3\mathbf{b}_3)$, and h_1, h_2, h_3 are integers.

These are well-known conditions for selective reflections (Laue spots and Bragg reflections, respectively).

We must now consider σ_2, which represents the extra scattering.

The arguments $w_{kx}, \ldots$ of quadratic form (5.8) are to be replaced by the components $w^l_{kx}, \ldots$ of the vectors

$$\mathbf{w}^l_k = \frac{f_k}{m_k^{\frac{1}{2}}} e^{i\mathbf{Q}.(\mathbf{r}_k+\mathbf{r}^l)}\mathbf{Q} = \mathbf{w}_k e^{i\mathbf{Q}.\mathbf{r}^l}, \qquad \ldots\ldots(6.12)$$

where $\mathbf{w}_k$ has now the same meaning (given in (5.9)) as before.

The scattering matrix $\mathscr{D}$ has now the elements

$$\mathscr{D}_{\substack{l-l' \\ kk' \\ xy}} = [\phi(\mathscr{D})]_{\substack{l-l' \\ kk' \\ xy}}, \qquad \ldots\ldots(6.13)$$

and (5.8) becomes $\sigma_2 = \sigma_0 \sum_{kk'} \sum_{xy} \mathscr{D}_{\substack{l-l' \\ kk' \\ xy}} w_{kx}{}^l w^*_{k'y}{}^{l'} e^{i\mathbf{Q}.(\mathbf{r}-\mathbf{r}')}$

As $\mathfrak{r}^{l'} - \mathfrak{r}^{l} = \mathfrak{r}^{l-l'}$ this can be written

$$\sigma_2 = \sigma_0 N \sum_{kx} \sum_{k'y} \mathscr{S}^{kk'}_{xy}(\mathbf{Q})\, w_{kx} w^*_{k'y} = \sigma_0 N (w^+ \mathscr{S}(\mathbf{Q}) w), \qquad \ldots\ldots(6.14)$$

where

$$\mathscr{S}^{kk'}_{xy}(\mathbf{Q}) = \sum_l \mathscr{S}^{kk'}_{lxy} e^{i\mathbf{Q}\cdot\mathfrak{r}^l}. \qquad \ldots\ldots(6.15)$$

(6.14) has exactly the same form as (5.8) apart from the fact that the coefficients depend on $\mathbf{Q}$.

Using the notation of submatrices, as in (6.3), (6.15) can be written

$$\mathscr{S}(\mathbf{Q}) = \sum_l \mathscr{S}_l e^{i(Ql)}, \quad (Ql) = Q_1 l_1 + Q_2 l_2 + Q_3 l_3. \qquad \ldots\ldots(6.16)$$

$\mathscr{S}(\mathbf{Q})$ is the Fourier transform of $\mathscr{S}_l$; the inverse transformation is

$$\mathscr{S}_l = \frac{1}{(2\pi)^3}\int \mathscr{S}(\mathbf{Q}) e^{-i(Ql)} d\overline{Q} = \frac{|a|}{(2\pi)^3}\int \mathscr{S}(\mathbf{Q}) e^{-i(\mathbf{Q}\cdot\mathfrak{r}^l)} d\mathbf{Q}, \qquad \ldots\ldots(6.17)$$

where $d\overline{Q} = dQ_1 dQ_2 dQ_3$, $d\mathbf{Q} = dQ_x dQ_y dQ_z$; $|a| = |a_1 a_2 a_3|$ is the volume of the cell and the integral is to be extended over a cell ($0 \leqslant Q_\alpha \leqslant 2\pi$).

We use now the well-known fact that a cyclic matrix (6.3) can be transformed into a diagonal matrix with the help of the unitary matrix

$$U(Q, l) = \frac{e^{i(Ql)}}{(2\pi)^{3/2}} 1,$$

where 1 is the unit matrix in the $3n$-dimensional subspace.

The element of the transformed matrix is

$$\begin{aligned}\mathscr{D}(Q, Q') &= \sum_{ll'} U(Q, l)\, \mathscr{D}_{l-l'} U(l', Q') \\ &= \sum_{l'} \mathscr{D}_{l'} e^{i(Q'l')} \sum_l (2\pi)^{-3} e^{i(lQ - Q'l)}.\end{aligned}$$

The last sum is a δ-function: for one has

$$\int f(Q') \sum_l (2\pi)^{-3} e^{i(lQ - Q'l)} d\overline{Q}' = \sum_l e^{i(Ql)} \int f(Q') e^{-i(Q'l)} \frac{d\overline{Q}'}{(2\pi)^3} = f(Q),$$

according to Fourier's theorem. Hence

$$\mathscr{D}(Q, Q') = \mathscr{D}(\mathbf{Q})\delta(Q - Q'), \qquad \ldots\ldots(6.18)$$

where $\mathscr{D}(\mathbf{Q})$ is defined in a way analogous to (6.16).

The transformed dynamical matrix is therefore diagonal in the $3n$-dimensional submatrices $\mathscr{D}(\mathbf{Q})$. Hence the transformed equations of motion have the form (4.4), where $\mathscr{D}$ is to be replaced by $\mathscr{D}(\mathbf{Q})$, and the eigen solutions satisfy (4.6), where the $\mathbf{e}^j_k$, ω_j are now also functions of $\mathbf{Q}$.

Further, it follows that the matrix equation (4.10) holds for each of these submatrices $\mathscr{D}(\mathbf{Q})$, and that the scattering matrix is also diagonal in the submatrices $\mathscr{S}(\mathbf{Q})$ given by this equation.

From (6.17) it is seen that the quantities $\mathscr{S}_0{}^{kk'}_{xy}$ are the elements of the matrix

$$\mathscr{S}_0 = \frac{1}{(2\pi)^3}\int \mathscr{S}(\mathbf{Q}) d\overline{Q}. \qquad \ldots\ldots(6.19)$$

Hence (6.4) becomes $M_k = \frac{1}{2m_k}\sum_{xy} Q_x Q_y \frac{1}{(2\pi)^3}\int \mathscr{S}^{kk}_{xy}(\mathbf{q}) d\bar{q}$. $\qquad \ldots\ldots(6.20)$

Thus all quantities occurring in σ_2 are expressed in terms of the matrix $\mathscr{S}(\mathbf{Q})$

The *result* can be *summarized* as follows:

The formulae for the scattering of a lattice are essentially the same as for a molecule formed by the atoms of the base cell, with the sole difference that the scattering matrix is a periodic function of the vector $\mathbf{Q}=\mathbf{K}'-\mathbf{K}$. If the elements of the dynamical matrix $\mathscr{D}$ are given, those of the scattering matrix $\mathscr{S}$ are found by the following steps:

$$\left.\begin{aligned}\mathscr{D}(\mathbf{Q}) &= \sum_l \mathscr{D}_l e^{i(\mathbf{Q}\mathbf{l})},\\ \mathscr{S}(\mathbf{Q}) &= \phi(\mathscr{D}(\mathbf{Q})) = \frac{\hbar}{2}\mathscr{D}^{-\frac{1}{2}}(\mathbf{Q})\coth\left(\frac{\hbar\mathscr{D}^{\frac{1}{2}}(\mathbf{Q})}{2kT}\right),\\ &= kT\left\{\mathscr{D}^{-1}(\mathbf{Q})+\frac{1}{12}\left(\frac{\hbar}{kT}\right)^2-\frac{1}{720}\left(\frac{\hbar}{kT}\right)^4\mathscr{D}(\mathbf{Q})+\ldots.\right\}.\end{aligned}\right\}\quad\ldots\ldots(6.21)$$

It is remarkable that this formula, written as a power series for the elements of the matrices, is already contained in Waller's principal paper (Dissertation, 1925).

If, vice versa, $\mathscr{S}(\mathbf{Q})$ is determined by observation of the scattering (which is possible at least in principle, see § 13), one can find the dynamical matrix $\mathscr{D}_l$; let $x=\psi(y)$ be the inverse function of $y=\phi(x)$; then

$$\mathscr{D}(\mathbf{Q})=\psi(\mathscr{S}(\mathbf{Q}))=kT\mathscr{S}^{-1}(\mathbf{Q})\left\{1+\frac{1}{12}\frac{\hbar^2}{kT}\mathscr{S}^{-1}(\mathbf{Q})+\frac{1}{180}\frac{\hbar^4}{k^2T^2}\mathscr{S}^{-2}(\mathbf{Q})+\ldots.\right\},$$

$$\mathscr{D}_l=\frac{1}{(2\pi)^3}\int\mathscr{D}(\mathbf{Q})e^{-i(\mathbf{Q}\mathbf{l})}d\bar{Q}.\qquad\ldots\ldots(6.22)$$

The law of the coefficients of the series has been determined (Born, 1942a).

The construction of $\mathscr{D}(\mathbf{Q})$ can be illustrated by the example of a monatomic linear chain in which only first neighbours act on one another. The matrix $\mathscr{D}$ is given by (6.3 *b*); its elements are numbers,

$$\mathscr{D}_1=\mathscr{D}_{-1}=-\tfrac{1}{2}\mathscr{D}_0=-\frac{\alpha}{m},\quad\mathscr{D}_2=\mathscr{D}_{-2}=\mathscr{D}_3=\mathscr{D}_{-3}=\ldots=0;$$

therefore $\mathscr{D}(Q)$ is also a number, namely,

$$\mathscr{D}(Q)=\mathscr{D}_0+\mathscr{D}_1e^{iQ}+\mathscr{D}_{-1}e^{-iQ}\qquad\ldots\ldots(6.23\,a)$$

$$=2\frac{\alpha}{m}(1-\cos Q)=4\frac{\alpha}{m}\sin^2\tfrac{1}{2}Q;$$

this leads to the well-known expression

$$\omega=(\mathscr{D}(Q))^{\frac{1}{2}}=2\left(\frac{\alpha}{m}\right)^{\frac{1}{2}}\sin\tfrac{1}{2}Q.\qquad\ldots\ldots(6.23\,b)$$

For the diatomic linear chain one has the matrices (6.3 *c*); hence

$$\mathscr{D}(Q)=\begin{bmatrix}\frac{2\alpha}{m_1}, & \frac{-\alpha}{(m_1m_2)^{\frac{1}{2}}}(1+e^{iQ})\\ \frac{-\alpha}{(m_1m_2)^{\frac{1}{2}}}(1+e^{-iQ}), & \frac{2\alpha}{m_2}\end{bmatrix}=\frac{\alpha}{(m_1m_2)^{\frac{1}{2}}}\begin{bmatrix}2\left(\frac{m_2}{m_1}\right)^{\frac{1}{2}}, & -(1+e^{iQ})\\ -(1+e^{-iQ}), & 2\left(\frac{m_1}{m_2}\right)^{\frac{1}{2}}\end{bmatrix}.$$

$\ldots\ldots(6.23\,c)$

The frequency spectrum is obtained from the corresponding secular determinant, adding ω^2 in the diagonal; the resulting quadratic equation can be written

$$\omega^4-\omega^2 2\alpha\left(\frac{1}{m_1}+\frac{1}{m_2}\right)+\frac{4\alpha^2}{m_1m_2}\sin^2\tfrac{1}{2}Q=0,\qquad\ldots\ldots(6.23\,d)$$

and leads to the well-known figure representing the two branches of the spectrum (see, for example, Born and Goeppert-Mayer, 1933, or Born, 1942 b).

The reciprocal matrix of $\mathscr{D}(Q)$ is

$$\mathscr{D}^{-1}(Q)=\frac{(m_1m_2)^{\frac{1}{2}}}{2\alpha(1-\cos Q)}\begin{pmatrix} 2\left(\frac{m_1}{m_2}\right)^{\frac{1}{2}} & 1+\varepsilon^{iQ} \\ 1+e^{-iQ} & 2\left(\frac{m_2}{m_1}\right)^{\frac{1}{2}}\end{pmatrix} \qquad \ldots\ldots (6.23\,e)$$

Hence according to (6.21) the scattering matrix is

$$\mathscr{S}(Q)=\begin{pmatrix}\mathscr{S}_{11}(Q) & \mathscr{S}_{12}(Q)\\ \mathscr{S}_{21}(Q) & \mathscr{S}_{22}(Q)\end{pmatrix}. \qquad \ldots\ldots (6.23\,f)$$

where

$$\left.\begin{aligned}\mathscr{S}_{11}(Q)&=\frac{m_1\boldsymbol{k}T}{\alpha(1-\cos Q)}-\frac{\hbar^2}{12\boldsymbol{k}T}-\frac{\alpha\hbar^4}{360m_1\boldsymbol{k}^3T^3}-\ldots\ldots\\ \mathscr{S}_{12}=\mathscr{S}^*_{21}(Q)&=(1+\varepsilon^{iQ})\left(\frac{(m_1m_2)^{\frac{1}{2}}\boldsymbol{k}T}{2\alpha(1-\cos Q)}-\frac{\alpha\hbar^4}{120(m_1m_2)^{\frac{1}{2}}\boldsymbol{k}^3T^3}-\ldots\right),\\ \mathscr{S}_{22}(Q)&=\frac{m_2\boldsymbol{k}T}{\alpha(1-\cos Q)}-\frac{\hbar^2}{12\boldsymbol{k}T}-\frac{\alpha\hbar^4}{360m_2\boldsymbol{k}^3T^3}-\ldots\ldots\end{aligned}\right\} \qquad \ldots\ldots (6.23\,g)$$

The simplest 3-dimensional model is the monatomic face-centred cubic lattice for which the matrix $\mathscr{D}$ is given in (6.3 *e*). From this one obtains

$$\left.\begin{aligned}&\mathscr{D}_{xx}(\mathbf{Q})=\frac{4}{m}\{\alpha(1-\cos Q_x\cos Q_y)+\beta(2-\cos Q_z\cos Q_x-\cos Q_x\cos Q_y)\},\\ &\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\\ &\mathscr{D}_{yz}(\mathbf{Q})=\frac{4}{m}\gamma\sin Q_y\sin Q_z\\ &\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\end{aligned}\right\} \qquad \ldots\ldots (6.23\,h)$$

This expression must go over into the elements of the matrix for molar elastic waves for small $Q_1=aQ_x, \ldots\ldots$ (a=lattice constant); by comparison with the latter (see (9.2), page 315) one can determine the three constants α, β, γ, in terms of Voigt's elastic constants c_{11}, c_{12}, c_{44}; if $2a$ is the edge of the elementary cube containing four atoms, the density is $m/2a^3$, and

$$\alpha=\frac{a}{2}(2c_{44}-c_{11}),\quad \beta=\frac{a}{2}c_{11},\quad \gamma=\frac{a}{2}(c_{12}+c_{44}). \qquad \ldots\ldots (6.23\,i)$$

For central forces one has $\alpha=0$, $\beta=\gamma$, which implies the Cauchy relation $c_{12}=c_{44}$, and, moreover, $c_{11}=2c_{44}$.

The reciprocal of (6.23 *h*) can easily be formed, and has been used by Born and Begbie (1943) for the discussion of the background scattering (see §13). Zachariasen (1940, 41) has used similar expressions for the simple cubic lattice, which is, however, not stable under central forces.

The formulae for the linear case (6.23 *f, g*) show explicitly that the main term of each element of $\mathscr{S}(Q)$ has, in the denominator, the factor $1-\cos Q=2\sin^2\frac{1}{2}Q$, which vanishes for $Q=0, 2\pi, 4\pi, \ldots\ldots$, i.e. in the points of the (linear) reciprocal lattice. Hence the scattering power is condensed around these lattice points. This holds as well in the three-dimensional case; for, according to (4.7), one has the expansion

$$\mathscr{S}(\mathbf{Q})=\sum_j \bar{e}^j(\mathbf{Q})\phi(\omega_j^2(\mathbf{Q}))e^j(\mathbf{Q})=\boldsymbol{k}T\sum_j\frac{\bar{e}^j(\mathbf{Q})e^j(\mathbf{Q})}{\omega_j(\mathbf{Q})^2}+\ldots. \qquad \ldots\ldots (6.24)$$

where $j=1, 2, \ldots\ldots 3N$ distinguishes the different branches of the elastic spectrum. Amongst these are three, the "acoustic branches", for which $\omega_j(\mathbf{Q})\to 0$ if $\mathbf{Q}$ approaches a reciprocal lattice point, while $e^j(\mathbf{Q})$ remains finite

Hence these acoustical branches are responsible for the infinite increase of scattering power in the vicinity of the reciprocal lattice points.

This is one of the results of lattice dynamics which Raman rejects. He says in his first publications that the optical branches produce the "modified scattering", and in his latest papers even contends that there are no optical branches (which would contain a practically infinite number of vibrations), but only a few single vibrations which modify the incident x rays in a peculiar way.

Here is the place to comment shortly on Raman's theory (already mentioned in § 1). His objections, at first only directed against the "thermal" theory of the extra spots, have been extended against the whole of lattice dynamics. His main arguments are empirical and are based on spectroscopic facts, namely, the small number of absorption and Raman lines observed in crystals. Raman considers these as a proof that there is no quasi-continuous infra-red spectrum of vibrations in the crystal, but only a few characteristic frequencies. His pupils have tried to support this assumption by calculating the specific heat from the sum of a few Einstein functions. However, lattice dynamics has given a full explanation of the small number of lines in the Raman and absorption spectrum by establishing certain selection rules, according to which most of the innumerable vibrations of the crystal are inactive with respect to absorption and scattering; for the secondary waves are extinct by interference in consequence of the random distribution of phases over the lattice vibrations. Concerning the specific heat, modern lattice theory has shown (Blackman, 1935a–42; Kellermann, 1942) that Debye's approximation, though moderately satisfactory over a large range of temperature, does not represent low temperature measurements with a high degree of accuracy, but that the exact formula (Born and v. Kármán, 1912, 13) accounts very well for the facts without introducing any arbitrary constant—in any case much better than Raman's formulae, which contain several *ad hoc* invented frequencies. Raman gives also a theoretical argument for his rejection of lattice theory, claiming that the method of cyclic boundary conditions is wrong. In order to refute this objection the case of a finite diatomic chain has been studied (Born, 1942b), and it has been shown that the correct solution leads again to the frequency equation given above for the infinite chain, (6.23 *d*), with a uniform distribution of frequencies over the Q scale.

Recently Ledermann (1943) has succeeded in giving a rigorous mathematical proof that the cyclic condition leads also in the three-dimensional case to the correct asymptotic expression for the distribution of frequencies. Empirically it is well supported by the fact that modern quantum theory of electrons in metals, which is completely based on the cyclic method, has produced a great number of results in perfect agreement with observation.

For the special problem of x-ray scattering the situation is still more favourable. For the cyclic method is used only for determining the density of the frequencies, which has to be known for performing summations over the whole spectrum, as e.g. in the theory of specific heats. But the distribution of scattering power of x rays is, in the main, independent of the distribution of the frequencies. It depends directly on the quasi-elastic forces, represented in the reciprocal space (by its Fourier transform), and this has nothing to do with the cyclic condition. Ledermann (1943) has given a rigorous proof that the results

obtained with the method of Fourier transformation, in particular the formulae (6.18), (6.21), are also asymptotically correct for a finite lattice. Raman's objections are therefore quite meaningless. His own explanation of the extra spots (see Mrs. Lonsdale's *Report*, page 281) consists in the assumption that the incident X rays excite the optical (infra-red) vibrations of the crystal and are then reflected by them with modified frequency. This is quite right, in so far as these excitations are contained among the quantum processes in formula (2.6); but they form only an extremely small part of the whole. The formulae used by Raman for describing this scattering mechanism are not based on quantum mechanics, but on classical wave theory.

§ 7. PHYSICAL INTERPRETATION

The derivation given here shows that the influence of the thermal agitation on the scattering can be described without reference to geometrical features of the waves in the crystal or to the frequencies of these vibrations. What matters are *waves of force*, represented by $\mathscr{G}(\mathbf{Q})$. One can consider $\mathbf{Q}$ as the wave vector of an elastic wave in the lattice; if Λ is its wave-length, one has, using (2.2),

$$\frac{1}{2\pi}|\mathbf{Q}| = \frac{1}{\Lambda} = \frac{2}{\lambda}\sin\tfrac{1}{2}\chi, \quad \lambda = 2\Lambda\sin\tfrac{1}{2}\chi, \qquad \ldots\ldots(7.1)$$

which is Bragg's interference condition for the reflection of light waves (λ) from a set of elastic waves (Λ) at the glancing angle $\frac{1}{2}\chi$. From the standpoint of quantum theory the definition (2.2) of $\mathbf{Q}$ can be interpreted as the momentum-conservation law

$$\mathbf{K}' = \mathbf{Q} + \mathbf{K}, \qquad \ldots\ldots(7.2)$$

expressing the condition that each X ray, scattered by thermal motion, is the result of the collision of a photon with momentum $\hbar\mathbf{K}$, and an elastic quantum with momentum $\hbar\mathbf{Q}$, producing a photon with momentum $\hbar\mathbf{K}'$ equal to the vector sum of the primary momenta.

This three-term collision equation (7.2) differs from the corresponding equation describing the Compton effect as a collision of a photon and a free electron; there one has a four-term energy and momentum equation

$$\omega + \Omega = \omega' + \Omega', \quad \mathbf{k} + \mathbf{K} = \mathbf{k}' + \mathbf{K}', \qquad \ldots\ldots(7.3)$$

where ω, $\mathbf{k}$ are the frequency and wave vector of the electron, Ω, $\mathbf{K}$ those of the photon before the collision, and the dashed symbols those after the collision. The reason for this difference is, that there are no free particles connected with sound waves ("phonons" say) in the same way as electrons are connected with de Broglie waves. The process is not analogous to the Compton effect, but to the scattering by bound electrons (Raman effect, in the general meaning of the term); it consists in double quantum transitions leading from one vibrational state to another. This is evident for an ordinary molecule, and nothing is changed by going over to a gigantic periodic molecule, a crystal. The fact that the normal modes here are plane waves is a secondary feature.

The intensity formula (6.14) represents the yield of the collisions, and this depends on the intensity of the elastic wave $\mathbf{Q}$, not measured geometrically (in amplitude), however, but dynamically by the elements of $S(\mathbf{Q}) = \phi(D(\mathbf{Q}))$.

Lonsdale and Smith (1941), Jahn (1941, 42) and, quite independently, Weigle (1941, 42), have suggested the following geometrical interpretation: The vector $\mathbf{Q}/2\pi$ is considered as the position vector in the *reciprocal space* in which the *reciprocal lattice* can be constructed from the fundamental vectors $\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3$; the lattice points $\mathbf{Q}^{(h)}/2\pi = (h\mathbf{b})$ according to (6.11) (corresponding to the selective reflection) are surrounded by a distribution σ_2 of scattering power. The vector $\mathbf{w}_k$ is then the reduced "reciprocal displacement" (dimensions: $[v] = m^{\frac{1}{2}}l$, $[w] = m^{-\frac{1}{2}}l^{-1}$). The elements of $S(\mathbf{Q})$ are periodic functions in the Q-space, but the scattering power itself is aperiodic, as the $\mathbf{w}_k$ are proportional to $\mathbf{Q}$ (5.9).

It is evident from the physical meaning (and will be shown below) that σ_2 is highly concentrated near the lattice points of the reciprocal lattice and weak in the intermediate regions. In order to investigate the vicinity of the lattice points, the relative position vector $\mathbf{q}$ is introduced by

$$\mathbf{Q} = 2\pi(h\mathbf{b}) + \mathbf{q} = \mathbf{Q}^{(h)} + \mathbf{q}. \qquad \ldots\ldots(7.4)$$

Then $\mathscr{S}(\mathbf{Q}) = \mathscr{S}(\mathbf{q})$, but $\mathbf{w}_k = \mathbf{w}_k^0 + \mathbf{w}_k^{(1)} + \ldots\ldots, \qquad \ldots\ldots(7.5)$

where $$\mathbf{w}_k^0 = \frac{f_k}{m_k^{\frac{1}{2}}} \mathbf{Q}^{(h)} e^{2\pi i(h\beta_k)}, \quad \beta_k = \mathbf{b} \,.\, \mathbf{r}_k, \qquad \ldots\ldots(7.6)$$

is constant, while $\mathbf{w}_k^{(1)}$ is linear in $\mathbf{q}$, etc.

Then $$\sigma_2(\mathbf{q}) = \sigma_2^0(\mathbf{q}) + \sigma_2^{(1)}(\mathbf{q}) + \ldots., \qquad \ldots\ldots(7.7)$$

where $$\left.\begin{aligned} \sigma_2^0(\mathbf{q}) &= \sigma_0 N(w^{0+}\mathscr{S}(\mathbf{q})w^0), \\ \sigma_2^{(1)}(\mathbf{q}) &= \sigma_0 N\{(w^{0+}\mathscr{S}(\mathbf{q})w^{(1)}) + (w^{(1)+}\mathscr{S}(\mathbf{q})w^0)\}. \\ &\ldots\ldots\ldots\ldots\ldots\ldots \end{aligned}\right\} \qquad \ldots\ldots(7.8)$$

In most cases the first approximation will suffice. Methods for its calculation will be explained in the following section. The distribution of $\sigma_2(\mathbf{q})$ around the corresponding lattice points can be represented by a set of *surfaces of isodiffusion* (Jahn, 1941, 42); $\sigma_2(\mathbf{q}) = \text{const}$. If these are constructed for all reciprocal

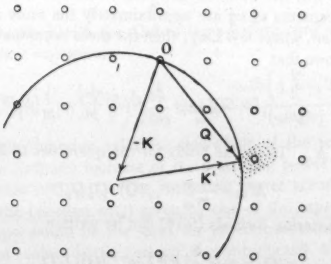


Figure 1. Ewald's construction for the sphere of propagation.

lattice points one has a representation of the complete function $\sigma_2(\mathbf{Q})$. In order to find the intensity for a given incident beam ($\mathbf{K}$) and a given direction of observation ($\mathbf{K}'$) one has to apply the equation (7.2). This can be done graphically with the help of *Ewald's sphere of propagation* (1913, 33). Draw the given

vector $-\mathbf{K}$ from the origin O of the reciprocal space and construct about its end-point C the sphere passing through O; as $|\mathbf{K}| = |\mathbf{K}'| = 2\pi/\lambda$, each point on this sphere defines a vector $\mathbf{Q}$ from O satisfying the equation $\mathbf{K}' - \mathbf{K} = \mathbf{Q}$ (see figure 1).

A selective reflection occurs whenever the sphere passes through a point $\mathbf{Q}^{(h)}$ of the reciprocal lattice.

The extra scattering is determined for any direction of $\mathbf{K}'$ by the value of $\sigma_2(\mathbf{Q})$ at the corresponding point of the sphere. If the orientation of the crystal relative to the incident ray is continuously changed from a position satisfying the Bragg condition the spot does not disappear instantaneously but gradually, becoming more and more diffuse.

The phenomena to be expected according to this construction are described in Mrs. Lonsdale's *Report*. These predictions depend on a knowledge of the scattering power $\sigma_2(\mathbf{Q})$; the methods for its calculation have now to be developed.

§ 8. APPROXIMATION METHOD FOR CALCULATING THE SCATTERING POWER

The property of crystal lattices used for simplifying the general expressions (Jahn, 1941, 42; Sarginson, 1942) is the existence of a set of frequencies which are very small. The frequency spectrum of a lattice $\omega_j(\mathbf{Q})$ consists of $3n$ branches; for three of these, $j = 1, 2, 3$, the "acoustical", one has

$$\omega_j(\mathbf{Q}) \to 0,$$

if $\mathbf{Q}$ approaches a point of the reciprocal lattice (selective reflection). For the rest, the $3(n-1)$ "optical" branches, $\omega_j(\mathbf{Q})$ is finite in the whole $\mathbf{Q}$ space.

If we now restrict our interest to the vicinity of the selective reflections, it suffices to consider only the acoustical branches; for $\phi(\omega^2) \to \infty$ when $\omega \to 0$, according to (4.8), (4.9). But for these branches the motion resembles that of an elastic continuum, the more so the longer the wave, i.e. the smaller $|\mathbf{Q}|$. Hence the displacements $\mathbf{e}_k^j/m_k^{\frac{1}{2}}$ are approximately the same for all particles of a cell, or $\mathbf{e}_k^j = \mathbf{e}^j m_k^{\frac{1}{2}}/m^{\frac{1}{2}}$, where $m = \sum_k m_k$; then the three vectors $\mathbf{e}^j$ are ortho-normal.

Then (4.7) shows that

$$\frac{\mathscr{S}_{xy}^{kk'}}{(m_k m_{k'})^{\frac{1}{2}}} = \frac{1}{(m_k m_{k'})^{\frac{1}{2}}}[\phi(\mathscr{D})]_{xy}^{kk'} \to \sum_{j=1}^{3} \phi(\omega_j^2)\frac{e_x^j e_y^j}{m} = \frac{1}{m}[\phi(\mathscr{D})]_{xy} = \frac{\mathscr{S}_{xy}}{m}, \qquad \ldots\ldots(8.1)$$

independent of k, k'. The same holds for the quantities (5.2):

$$M_{kk'} \to \frac{1}{m}\sum_{xy} \mathscr{S}_{xy}(\mathbf{Q}) \cdot Q_x Q_y. \qquad \ldots\ldots(8.2)$$

Finally, the scattering formula (6.14) can be written

$$\sigma_2(\mathbf{Q}) = \sigma_0 N \frac{1}{m}|F|^2 \sum_{xy} \mathscr{S}_{xy}(\mathbf{Q}) Q_x Q_y, \qquad \ldots\ldots(8.3)$$

where F is the same lattice structure factor (6.10) which enters into the expression for the Laue scattering.

Born and Sarginson (1941) have obtained an explicit formula for cubic lattices holding in the whole reciprocal space, by replacing the three acoustical frequencies by a single one.

Then one has (see (8.1))

$$\mathscr{D}_{xx}=\mathscr{D}_{yy}=\mathscr{D}_{zz}=\phi(\omega^2), \quad \mathscr{D}_{yz}=\mathscr{D}_{zx}=\mathscr{D}_{xy}=0;$$

hence (8.2) becomes with (2.2)

$$M_{kk'}\to\frac{1}{m}\phi(\omega^2)\mathbf{Q}^2=\frac{1}{m}\left(\frac{4\pi}{\lambda}\right)^2\sin^2\tfrac{1}{2}\chi\,.\,\phi(\omega^2),$$

and the scattering power

$$\sigma_s=\sigma_0 N\frac{1}{m}|F|^2\left(\frac{4\pi}{\lambda}\right)^2\sin^2\tfrac{1}{2}\chi\,.\,\phi(\omega^2).$$

Let ω_0 be the maximum frequency, and write

$$\omega=\omega_0 w(\mathbf{Q}), \qquad \ldots\ldots(8.4)$$

then $w(\mathbf{Q})$ is a dimensionless periodic function depending on the structure; the square of the frequency is given by the quasi-elastic forces, hence

$$w^2(\mathbf{Q})=\sum_l \Phi_l\sin^2\tfrac{1}{2}(l\mathbf{Q})=\tfrac{1}{2}\sum_l\Phi_l[1-\cos(l\mathbf{Q})], \quad \sum_l\Phi_l=1. \qquad \ldots\ldots(8.5)$$

For the three cubic Bravais lattices (s = simple, f = face-centred, b = body-centred) one obtains, under the assumption that only first neighbours act on one another ($\Phi_l=0$ for $l>1$):

$$\left.\begin{aligned}
&\text{(s)}: \ w^2(\mathbf{Q})=\tfrac{1}{3}\{1-\tfrac{1}{3}[\cos Q_1+\cos Q_2+\cos Q_3]\},\\
&\text{(f)}: \ w^2(\mathbf{Q})=\tfrac{1}{12}\{1-\tfrac{1}{3}[\cos(Q_2+Q_3)+\cos(Q_3+Q_1)+\cos(Q_1+Q_2)\\
&\qquad\qquad +\cos(Q_2-Q_3)+\cos(Q_3-Q_1)+\cos(Q_1-Q_2)]\},\\
&\text{(b)}: \ w^2(\mathbf{Q})=\tfrac{1}{4}\{1-\tfrac{1}{4}[\cos(Q_1+Q_2+Q_3)+\cos(-Q_1+Q_2+Q_3)\\
&\qquad\qquad +\cos(Q_1-Q_2+Q_3)+\cos(Q_1+Q_2-Q_3)]\}.
\end{aligned}\right\} \qquad \ldots\ldots(8.6)$$

If, now, the expression (4.8) for $\phi(\omega^2)$ is introduced there appears in the argument the factor

$$\frac{\hbar\omega_0}{k}=\Theta,$$

which can be identified with the Debye temperature, and one obtains for the intensity the explicit expression

$$\sigma_s=\sigma_0 N\frac{1}{m}|F|^2\left(\frac{4\pi}{\lambda}\right)^2\frac{\hbar^2}{2k\Theta}\sin^2\tfrac{1}{2}\chi\,\frac{\coth\left(\frac{1}{2}\frac{\Theta}{T}w(\mathbf{Q})\right)}{w(\mathbf{Q})}, \qquad \ldots\ldots(8.7)$$

which is a periodic function valid for the whole of the reciprocal space. It shows clearly the (infinite) increase of the scattering power near the reciprocal lattice points where $w(\mathbf{Q})\to 0$. The scattering power according to (8.7) is not isotropic, since the function $w(\mathbf{Q})$ given by one of the expressions (8.6) is not. But the anisotropy would be the same for all reciprocal lattice points and for all substances of cubic lattice, which is contradicted by experiment (see Mrs. Lonsdale's *Report*).

A rigorous expression for $\mathscr{D}=kT\mathscr{D}^{-1}(\mathbf{Q})$ holding in the whole reciprocal space can easily be constructed for cubic models, e.g. for the monatomic face-centred cubic lattice with the help of the formula (6.23 *h*) for $\mathscr{D}(\mathbf{Q})$. Zachariasen (1940) has used this method for the simple cubic lattice; in his applications, however, he considered only the vicinity of the reciprocal lattice

points by expanding the periodic functions. The result can be obtained without considering a model, and for any crystal, by the general argument that the acoustical branches near the reciprocal lattice points are approximately identical with the vibrations of an anisotropic continuous crystal. A wave of frequency ω and wave vector $\mathbf{Q}$ in an elastic crystal satisfies equations of the form (see Born, 1923)

$$\omega^2 e_x^j = \mathscr{D}_{xx} e_x^j + \mathscr{D}_{xy} e_y^j + \mathscr{D}_{xz} e_z^j, \ldots ; \qquad \ldots\ldots(8.8)$$

the coefficients are quadrics of the components of $\mathbf{Q}$, e.g.,

$$\rho \mathscr{D}_{xx} = c_{11} Q_x^2 + c_{66} Q_y^2 + c_{55} Q_z^2 + 2c_{65} Q_y Q_z + 2c_{51} Q_z Q_x + 2c_{16} Q_x Q_y, \quad \ldots\ldots(8.9)$$

where ρ is the density and $c_{\alpha\beta}$ Voigt's elastic constants. (The complete expressions can be found in any textbook of elasticity, e.g. Lamb, p. 287, or in the papers of Jahn, 1942 a, 42 b, and Born, 1942 a.)

In this approximation one has, according to (6.21),

$$\mathscr{S}(\mathbf{Q}) = \boldsymbol{k} T \mathscr{D}^{-1}(\mathbf{Q}); \qquad \ldots\ldots(8.10)$$

this gives expressions of the form

$$\mathscr{D}_{xx}^{-1}(\mathbf{Q}) = \frac{1}{\det(\mathscr{D})} \begin{vmatrix} \mathscr{D}_{yy} \mathscr{D}_{yz} \\ \mathscr{D}_{zy} \mathscr{D}_{zz} \end{vmatrix}, \ldots\ldots, \qquad \ldots\ldots(8.10\,a)$$

in which the values (8.9) have to be substituted. In this way the scattering matrix is expressed in terms of the elastic constants.

For the vicinity of a Laue spot the approximation (7.7) can be used.

The expression $$d(\mathbf{q}) = \sum_{xy} \mathscr{D}_{xy}^{-1}(\mathbf{q}) Q_x^{(h)} Q_y^{(h)} \qquad \ldots\ldots(8.11)$$

may be termed the "diffusion function", since it describes the diffusion of scattered intensity about a Laue spot in virtue of the thermal motion. The exponent M_k of f_k becomes independent of k, namely (see 6.20):

$$M_k \to M = T/2T_h, \qquad \ldots\ldots(8.12)$$

where $$\frac{1}{T_h} = \frac{\boldsymbol{k}}{(2\pi)^3 m} \int d(\mathbf{q}) d\bar{q} = \sum_{xy} Q_x^{(h)} Q_y^{(h)} \frac{\boldsymbol{k}}{(2\pi)^3 m} \int \mathscr{D}_{xy}^{-1}(\mathbf{q}) d\bar{q}, \qquad \ldots\ldots(8.13)$$

and the structure factor F, (6.10), can be written

$$F = F_0 e^{-M}, \quad F_0 = \sum_k f_k^0 e^{i\mathbf{Q} \cdot \mathbf{r}_k}. \qquad \ldots\ldots(8.14)$$

The scattering power in the neighbourhood of a Laue spot $\mathbf{Q}^{(h)}$ is now approximately $$\sigma_2 = \sigma_0 \frac{N}{m} |F_0|^2 \boldsymbol{k} T e^{-T/T_h} d(\mathbf{q}). \qquad \ldots\ldots(8.15)$$

Several authors (quoted in § 1, p. 296) have derived this formula with the expression for $d(\mathbf{q})$ in terms of the eigenvectors, namely,

$$d(\mathbf{q}) = \sum_{j=1}^{3} \frac{1}{\omega_j^2} (\mathbf{Q}^{(h)} \cdot \mathbf{e}^j)^2, \qquad \ldots\ldots(8.16)$$

which follows by applying (4.7), or (6.24), to $\mathscr{D}^{-1}$. This is, however, unsatisfactory, as ω_j, $\mathbf{e}^j$ are irrational quantities (ω_j^2 are the roots of a cubic equation), whereas $\mathscr{D}^{-1}$ can be directly obtained by rational operations.

§ 9. CUBIC CRYSTALS

In the case of a cubic lattice $\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3$ are orthogonal vectors of equal length a; $\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3$ parallel to them of length $\frac{1}{a}$. Hence $Q_1^{(h)} = aQ_x^{(h)} = 2\pi h_1, \ldots$ and

$$d(\mathbf{q}) = \frac{4\pi^2}{a^2}\{\mathscr{D}_{xx}^{-1}(\mathbf{q})h_1^2 + \ldots + 2\mathscr{D}_{yz}^{-1}(\mathbf{q})h_2h_3 + \ldots\}. \quad \ldots\ldots(9.1)$$

A cubic crystal has three independent elastic constants: in Voigt's notation, c_{11}, c_{12}, c_{44}. Then (8.9) reduces to

$$\left.\begin{aligned} \rho\mathscr{D}_{xx}(\mathbf{q}) &= c_{11}q_x^2 + c_{44}(q_y^2 + q_z^2), \ldots \\ \rho\mathscr{D}_{yz}(\mathbf{q}) &= (c_{12} + c_{44})q_yq_z, \ldots \end{aligned}\right\}. \quad \ldots\ldots(9.2)$$

The elastic anisotropy is measured by the constant

$$c_{11} - c_{12} - 2c_{44} = \epsilon \quad \ldots\ldots(9.3)$$

which vanishes for elastic isotropic bodies.

The determinant of the coefficients of (9.2), expressed in terms of $q_1 = aq_x, \ldots, q^2 = q_1^2 + q_2^2 + q_3^2$, and multiplied by ρ^3q^6, is

$$\begin{aligned} A_0 = {} & c_{11}c_{44}^2q^6 + \epsilon c_{44}(c_{11} + c_{44})q^2(q_2^2q_3^2 + q_3^2q_1^2 + q_1^2q_2^2) \\ & + \epsilon^2(c_{11} + 2c_{12} + c_{44})q_1^2q_2^2q_3^2. \end{aligned} \quad \ldots\ldots(9.4)$$

Then
$$\left.\begin{aligned} \mathscr{D}_{xx}^{-1}(\mathbf{q}) &= \frac{a^2\rho}{A_0}\{c_{44}q^2[c_{44}q_1^2 + c_{11}(q_2^2 + q_3^2)] + \epsilon(c_{11} + c_{12})q_2^2q_3^2\}, \\ &\ldots\ldots \\ \mathscr{D}_{yz}^{-1}(\mathbf{q}) &= -\frac{a^2\rho}{A_0}(c_{12} + c_{44})q_2q_3(c_{44}q^2 + \epsilon q_1^2). \\ &\ldots\ldots \end{aligned}\right\} \quad \ldots\ldots(9.5)$$

If this is introduced in (9.1) the diffusion function is given by

$$d(\mathbf{q}) = 4\pi\rho\frac{A(\mathbf{q})}{A_0(\mathbf{q})}, \quad \ldots\ldots(9.6)$$

where
$$\begin{aligned} A(\mathbf{q}) = {} & h_1^2\{c_{44}q^2[c_{44}q_1^2 + c_{11}(q_2^2 + q_3^2)] + \epsilon(c_{11} + c_{12})q_2^2q_3^2\} + \ldots \\ & - 2h_2h_3(c_{12} + c_{44})q_2q_3(c_{44}q^2 + \epsilon q_1^2) - \ldots \end{aligned} \quad \ldots\ldots(9.7)$$

These formulae are already contained in Waller's Dissertation (1925); they have been applied to the observation of extra spots by Jahn (1942) and Sarginson (1942).

§ 10. INFLUENCE OF TEMPERATURE

The influence of temperature on the Laue spots and on the extra spots is very different. On lowering the temperature the Laue spots become stronger and sharper, while the extra spots become weaker, smaller, and finally disappear. Hence the theoretical treatment must be different in both cases; for the Laue scattering, one needs a formula holding down to the absolute zero, while for the extra spots approximations valid for higher temperatures will temporarily suffice.

The solution of the first problem is due to Debye (1913). The temperature enters the expression σ_1, (6.9), only through $|F|^2$, where F is given by (6.10);

it is a rather complicated function of T, as M_k is different for each kind of atom. Debye's approximation is the same as that used by him in the theory of specific heat: he neglects the optical branches of the spectrum and treats the acoustical vibrations as if they were produced by an elastic continuum. This amounts to a simple generalization of the formulae (8.12), (8.13): M_k is replaced by M, so that F is given by (8.14), and the intensity, (6.9), becomes

$$\sigma_1 \propto e^{-2M}; \qquad \ldots\ldots(10.1)$$

but instead of (8.12), (8.13), one has

$$M = \tfrac{1}{2}\sum_{xy} Q_x^{(h)} Q_y^{(h)} \frac{1}{(2\pi)^3 m} \int \mathscr{S}_{xy}(\mathfrak{q})\, d\bar{q},$$

where

$$\mathscr{S}_{xy} = [\phi(\mathscr{D})]_{xy} = \sum_j \phi(\omega_j^2) e_x^j e_y^j.$$

For cubic crystals this can be approximately replaced by

$$\mathscr{S}_{xy} = \tfrac{1}{3}\sum_j \phi(\omega_j^2) \sum_j e_x^j e_y^j = \tfrac{1}{3}\sum_j \phi(\omega_j^2)\delta_{xy};$$

hence with (6.11), $\quad Q_x^{(h)} = 2\pi h_1/a, \ldots.$

$$2M = \frac{1}{m}\left(\frac{2\pi}{a}\right)^2 (h_1^2 + h_2^2 + h_3^2) \frac{1}{(2\pi)^3} \frac{1}{3} \sum_{j=1}^{3} \int \phi(\omega_j^2)\, d\bar{q}.$$

Each of the three integrals over the cube ($0 < q_\alpha < 2\pi$) is approximately replaced by that over a sphere of equal volume, radius q_m (element of solid angle $d\Omega$):

$$d\bar{q} = dq_1\, dq_2\, dq_3 = q^2\, dq\, d\Omega, \quad (2\pi)^3 = \frac{4\pi}{3} q_m^3, \quad q_m = (6\pi^2)^{\frac{1}{3}}.$$

The connection between q and ω is supposed to be that of elastic waves in a continuous medium, $q = \dfrac{d\omega}{c}$, where c is the velocity of sound, a function of the direction of the wave. Then

$$2M = \frac{2a}{m}(h_1^2 + h_2^2 + h_3^2) \frac{1}{c_D^3} \int_0^{\omega_m} \phi(\omega^2)\omega^2\, d\omega, \quad \omega_m = c_D q_m/a = \frac{c_D}{a}(6\pi^2)^{\frac{1}{3}}.$$

Debye's mean velocity of sound, c_D, is the average for all directions of the weighted longitudinal and transversal waves:

$$\frac{1}{c_D^3} = \frac{1}{3}\iint \left(\frac{2}{c_t^3} + \frac{1}{c_l^3}\right) \frac{d\Omega}{4\pi}. \qquad \ldots\ldots(10.2)$$

If, now, $\phi(\omega^2)$ is substituted from (4.8) and ω replaced by

$$\xi = \frac{\hbar\omega}{kT},$$

then a simple calculation gives

$$2M = \frac{3\hbar^2}{mk\Theta}\left(\frac{2\pi}{a}\right)^2 (h_1^2 + h_2^2 + h_3^2)\, \Psi\left(\frac{\Theta}{T}\right), \qquad \ldots\ldots(10.3)$$

where (with $a = (V/N)^{\frac{1}{3}}$)

$$\Theta = \frac{\hbar\omega_m}{k} = \frac{\hbar c_D}{ak}(6\pi^2)^{\frac{1}{3}} = \frac{(2\pi\hbar)c_D}{k}\left(\frac{3N}{4\pi V}\right), \qquad \ldots\ldots(10.4)$$

$$\Psi(x) = \frac{1}{x^3}\int_0^x \tfrac{1}{2}\xi \coth(\tfrac{1}{2}\xi)\, d\xi. \qquad \ldots\ldots(10.5)$$

(10.3) is the formula of Debye (1914), corrected by a factor 2, due to Waller (see § 1). It neglects all details of lattice structure, but has the advantage of holding down to zero temperature, i.e. $x=\Theta/T\to\infty$, where $\Psi(x)$ approaches a constant value (1/4). This is important in connection with the question of zero point energy, the existence of which could indeed be demonstrated with the help of this effect (see e.g. Laue's book, 1941).

This theory has been generalized for polyatomic and non-cubic crystals by considering the constant Θ in the temperature factor (10.3) as different for different kinds of atom and even for different directions of vibrations; see the detailed discussion in Mrs. Lonsdale's *Report*, p. 263.

In the expression (6.14) for the extra scattering the temperature enters not only through the Debye factor, $\exp(-2M)$, but also directly in the scattering matrix $\mathscr{S}=\phi(\mathscr{D})$, as ϕ depends on T. For low temperatures one has

$$\mathscr{S}=\hbar\mathscr{D}^{-1}(\tfrac{1}{2}+\exp(-\hbar\mathscr{D}/kT)+\ldots),$$

and the problem of determining the elements of this matrix in terms of those of $\mathscr{D}$ would not be easy. But it is fortunately unnecessary, for, as noted above, the diffuse scattering becomes weaker and finally unobservable with decreasing temperature.

Hence the interest of the problem lies principally in the domain of higher temperatures, $T>\Theta$. Here the approximation of §§ 8 and 9 can be used, and $d(q)$ is independent of temperature if the small variation of the elastic constants with T is neglected. (8.15) shows that for a crystal of any symmetry

$$\sigma_{\mathrm{s}}\propto Te^{-T/T_{\mathrm{h}}}, \qquad \ldots\ldots(10.6)$$

where T_{h} is given by (8.13)

T_{h} may be called the "inversion temperature", for the function (10.6) increases with T until $T=T_{\mathrm{h}}$ where it has a maximum. It is important to see where this inversion is to be expected. It is, however, doubtful whether the theory as developed here is still valid up to $T=T_{\mathrm{h}}$. For then $2M=T/T_{\mathrm{h}}$ becomes unity; the quantities $M_{kk'}$, (5.2), are in these circumstances possibly not small in the whole reciprocal space, and the expansion with respect to them, on which this theory—and all theories published so far, with the exception of that of Weigle and Bleuler (1942) mentioned on page 302—is based, may become invalid. It is, however, likely that the correct theory (contained in the formulae of §§ 2, 3, 4 and 5, but not developed) will also lead to an inversion. In any case, T_{h} must be estimated. This can be done for cubic crystals either with the help of (8.13), (9.6) or by Debye's approximation. For $x\ll 1$ one has $\Psi(x)\sim\frac{1}{4}x$; hence from (8.12) and (10.3)

$$T_{\mathrm{h}}=\frac{T_0}{h_1^2+h_2^2+h_3^2}, \quad T_0=\frac{mka^2\Theta^2}{12\pi^2\hbar^2}=\frac{\Theta}{12\pi^2}\left(\frac{a}{l}\right)^2, \qquad \ldots\ldots(10.7)$$

where (μ = atomic mass in atomic units)

$$l=\frac{\hbar}{(mk\Theta)^{\frac{1}{2}}}=\frac{6{\cdot}96\times10^{-8}}{(\mu\Theta)^{\frac{1}{2}}}\ \mathrm{cm}. \qquad \ldots\ldots(10.8)$$

This quantity has the dimension "length" and may be called "dynamical lattice constant"; it is the (root mean square) amplitude of a linear oscillator of frequency $\omega_0=k\Theta/\hbar$, if the energy of vibration is $k\Theta$:

$$ml^2\omega_0^2=\hbar\omega_0=k\Theta.$$

Table 1 contains Θ and T_0 for a few substances. One sees that the latter is fairly high, but for spots with high Miller indices T_h may be about room temperature. The theory, therefore, predicts a difference in the temperature dependence of low and high order spots. This is confirmed by Laval's (1939) observations.

Table 1

°K.	Face-centred			Body-centred		Pseudo-simple			Diamond
	Cu	Pb	Al	Li	Na	NaCl	KCl	KBr	C
Θ	315	88	398	510	202	281	230	177	2340
T_0	14 450	6 880	12 260	3 800	2 970	12 850	13 670	14 230	146 400

A detailed comparison of the temperature law (10.6) with the observations of Laval on KCl with Mo$K\alpha$ radiation has led to an excellent confirmation of the theory (Born and Lonsdale, 1942). An account of this work is given in Mrs. Lonsdale's *Report*. It proves conclusively the validity of the thermal theory.

For $T \leqslant \Theta$ it is necessary to consider also the influence of the application of quantum theory on the scattering matrix itself; the function $\mathscr{S} = \phi(\mathscr{P})$, (4.10), depends on T. The first correction consists in adding to $d(\mathbf{q})$

$$\frac{1}{2}\left(\frac{\hbar}{kT}\right)^2 (Q_x^2 + Q_y^2 + Q_z^2) = \frac{1}{2}\left(\frac{\hbar}{akT}\right)^2 (Q_1^2 + Q_2^2 + Q_3^2). \quad \ldots\ldots (10.9)$$

This clearly represents a structureless background which increases in intensity with increasing distance from the origin of the reciprocal space. Its value at the reciprocal lattice point h_1, h_2, h_3 is

$$\frac{1}{2}\left(\frac{2\pi\hbar}{akT}\right)^2 (h_1^2 + h_2^2 + h_3^2).$$

This has to be compared with $d(\mathbf{q})$ with respect to order of magnitude. For this purpose it suffices to take $\epsilon = 0$; then (as shown below (11.11)) $d(\mathbf{q})$ reduces to

$$d(\mathbf{q}) = \frac{4\pi\rho(h_1^2 + h_2^2 + h_3^2)}{\gamma q^2},$$

where γ is of the dimension of an elastic constant but depends on the direction of $\mathbf{q}$. If γ is replaced by a constant average and the velocity $(\rho/\gamma)^{\frac{1}{2}}$ is identified with Debye's c_D, (10.2), one has, using (10.4),

$$d(\mathbf{q}) = \frac{4\pi(h_1^2 + h_2^2 + h_3^2)}{c_D^2 q^2} = 4\pi(6\pi^2)^{\frac{1}{3}}\left(\frac{\hbar}{ak\Theta}\right)^2 \frac{h_1^2 + h_2^2 + h_3^2}{q^2}.$$

Hence the ratio of the correction term to $d(\mathbf{q})$ is approximately

$$\frac{1}{2(36\pi)^{\frac{1}{3}}}\left(\frac{\Theta q}{T}\right)^2 \sim \frac{1}{9}\left(\frac{\Theta q}{T}\right)^2,$$

and the radius of a spherical region about a reciprocal lattice point inside of which the scattering power is essentially stronger than the background is of the order

$$q \sim 3T/\Theta. \quad \ldots\ldots (10.10)$$

At a given temperature this radius is much smaller for diamond than for all other substances; hence the spots of diamond will be particularly sharp.

Other peculiarities of diamond are described in Mrs. Lonsdale's *Report*. The majority of diamond crystals show two types of extra spots, primary spots which are temperature sensitive and just of the type predicted by the thermal theory, and secondary spots which are hardly influenced by temperature but are structure sensitive. There exists a rarer type of diamond (so-called type II, while the normal crystals are called type I) which gives no traces of the secondary spots. It is obvious that the secondary spots are produced by irregularities of structure connected with internal stresses, and there are indications, indeed, that the crystals of type II are free from such stresses which exist in the ordinary crystals (type I). A systematic theory of this structural scattering does not exist, but its outlines are obvious; they may be sketched briefly as follows. The displacements of the atoms produced by the internal stresses can be represented, by a Fourier transformation, as a distribution in the reciprocal space which determines the scattering in the same way as in the case of thermal displacements. Vice versa, from the observed distribution of scattering power in the reciprocal space it should be possible to obtain information about the distribution of the actual displacements (strains) and stresses in ordinary space.

The material available at present is, however, hardly sufficient to give reliable results.

§ 11. ANALYTICAL EXPRESSIONS FOR THE POSITION OF INTENSITY MAXIMA

In order to find analytical expressions for the position of "extra spots", i.e. intensity maxima of the background radiation, one has to specify the experimental arrangements. In connection with calculations of W. H. Bragg

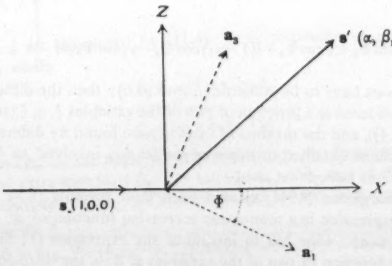


Figure 2.

(1941 a–41 c), on his geometrical theory the following case has been investigated (Born and Sarginson, 1941; Sarginson, 1942).

The x-axis is taken as the direction of the incident beam, i.e. $\mathfrak{s}$ (1, 0, 0). A cubic crystal is rotated about its $\mathfrak{a}_2$-axis, which coincides with the y-axis, by an angle Φ. Then (figure 2)

$$\mathfrak{a}_1 = (a\cos\Phi, 0, a\sin\Phi), \quad \mathfrak{a}_2 = (0, a, 0), \quad \mathfrak{a}_3 = (-a\sin\Phi, 0, a\cos\Phi).$$

Let α, β, γ be the components of $\mathfrak{s}'$,

$$\alpha^2+\beta^2+\gamma^2=1; \qquad \ldots\ldots(11.1)$$

then $$Q_x=\frac{2\pi}{\lambda}(\alpha-1), \quad Q_y=\frac{2\pi}{\lambda}\beta, \quad Q_z=\frac{2\pi}{\lambda}\gamma.$$

Substituting this in (6.6) one obtains

$$\left.\begin{aligned} Q_1&=\frac{2\pi a}{\lambda}\{(\alpha-1)\cos\Phi+\gamma\sin\Phi\},\\ Q_2&=\frac{2\pi a}{\lambda}\beta,\\ Q_3&=\frac{2\pi a}{\lambda}\{-(\alpha-1)\sin\Phi+\gamma\cos\Phi\}. \end{aligned}\right\} \qquad \ldots\ldots(11.2)$$

A selective reflection* corresponds to the values $Q_\alpha=2\pi h_\alpha$ $(\alpha=1, 2, 3)$; (11.1) and (11.2) are four equations from which the corresponding values of $\alpha, \beta, \gamma, \Phi$ can be obtained. Let these be α_h, β_h, γ_h, Φ_h. Now the crystal is turned through a small angle ϕ; let ξ, η, ζ be the corresponding small changes of the components of $\mathfrak{s}'$:

$$\phi=\Phi-\Phi_h, \quad \xi=\alpha-\alpha_h, \quad \eta=\beta-\beta_h, \quad \zeta=\gamma-\gamma_h \qquad \ldots\ldots(11.3)$$

It follows from (11.1) that in first approximation

$$\xi\alpha_h+\eta\beta_h+\zeta\gamma_h=0. \qquad \ldots\ldots(11.4)$$

The deviations q defined by (7.4) are obtained from (11.2):

$$\left.\begin{aligned} q_1&=\frac{2\pi a}{\lambda}\{\xi\cos\Phi_h+\zeta\sin\Phi_h+[(1-\alpha_h)\sin\Phi_h+\gamma_h\cos\Phi_h]\phi\},\\ q_2&=\frac{2\pi a}{\lambda}\eta,\\ q_3&=\frac{2\pi a}{\lambda}\{-\xi\sin\Phi_h+\zeta\cos\Phi_h+[(1-\alpha_h)\cos\Phi_h-\gamma_h\sin\Phi_h]\phi\}. \end{aligned}\right\} \qquad \ldots\ldots(11.5)$$

These expressions have to be substituted into (9.6); then the diffusion function $d(\mathbf{q})$ can be considered as a function of two of the variables ξ, η, ζ (since the third is given by (11.4)), and the maxima of intensity are found by differentiating.

The expressions obtained in this way are far too involved to be discussed, and simplifications have to be used.

Born and Sarginson (1941) started from their formula (8.7). As the last factor in this expression is a monotonic increasing function of w, maxima of σ_z are minima of $w(\mathbf{q})$. One has to introduce the expression (11.2) in (8.7) and consider w as a function of two of the variables α, β, γ, the third following from (11.1).

* The expression "selective reflection" is used in this *Report* in order to avoid the distinction between the Laue and the Bragg case. The wording of the text above corresponds to a Bragg reflection (λ given, Φ_h determined from the interference condition), hence $\alpha_h=\cos 2\vartheta_B$, where ϑ_B is the glancing angle of Bragg. One can, however, just as well consider Φ as given and determine $\alpha_h', \beta_h', \gamma_h', \lambda_h$ as functions of h_1, h_2, h_3, Φ; then $\alpha_h'=\cos 2\vartheta_L$, where ϑ_L is the glancing angle for a Laue spot, belonging to the wave-length λ_h.

This leads to an interesting connection with the "geometrical" theory suggested by Preston (1939 a, 39 b) and developed by W. H. Bragg (1941 a–41 c); see § 1.

Assuming that there are small groups of atoms in the crystal which scatter independently from the rest of the lattice, Bragg obtained, with the help of elementary interference considerations, expressions for the intensity which are simple periodic functions of Q_1, Q_2, Q_3; if one takes eight atoms at the corners of a cube, this function is just $1-2w^2(\mathbf{Q})$, where $w(\mathbf{Q})$ is the first of the expressions (8.6). Hence Bragg's "intensity" is also given by the condition that $w(\mathbf{Q})$ shall be a minimum. Now the assumption of the small independent groups of atoms is hardly reconcilable with our knowledge about lattice-dynamics. The reason why Bragg was quite successful in accounting for some simple features of the positions of the extra spot is now obvious; the form of $w(\mathbf{Q})$ used by Bragg is a consequence of the assumption that only first neighbours act on one another. One can also say that there are no geometrically independent groups of atoms, but dynamically coupled groups of atoms.

Born and Sarginson have discussed the formulae (8.6) for the neighbourhood of a Laue spot. If the expressions (8.6) are expanded in powers of $\mathbf{Q}-\mathbf{Q}^{(h)}=\mathbf{q}$, the first term is proportional to q^2. This assumption, which means complete isotropy of the scattering power about the reciprocal lattice points, was first made by Faxén (1923), who derived from it a formula for the position of the extra spot; the same result has been repeatedly published by other authors (Raman and Nilakantan, 1940; Jauncey, 1941; Preston, 1941; Lonsdale and Smith, 1941; Born and Sarginson, 1941). If $d(\mathbf{q})$ is replaced by a function of q^2 (which obviously must be monotonic and decreasing with q^2) then the maxima of intensity are the minima of

$$q^2=\left(\frac{2\pi a}{\lambda}\right)^2\{(\xi^2+\eta^2+\zeta^2)+2[\gamma_h\xi+(1-\alpha_h)\zeta]\phi+2[(1-\alpha_h)^2+\gamma_h^2]\phi^2,\} \qquad \ldots\ldots(11.6)$$

where ξ, η, ζ are connected by (11.4). The minimum of q^2 is obtained for $\xi=\xi_m, \ldots\ldots$ where

$$\left.\begin{aligned}\xi_m&=-(1-\alpha_h)\gamma_h\phi,\\ \eta_m&=\beta_h\gamma_h\phi,\\ \zeta_m&=(\alpha_h+\gamma_h^2-1)\phi.\end{aligned}\right\} \qquad \ldots\ldots(11.7)$$

Let $\chi_m=2\vartheta_m$ be the angle of deflection (from the direction of the incident beam) of the extra spot $(\alpha_h+\xi_m, \beta_h+\eta_m, \gamma_h+\zeta_m)$, $2\vartheta_B$ that of the corresponding Bragg reflection $(\alpha_h, \beta_h, \gamma_h)$. Then the length of the vector product of these two unit vectors is

$$[(\beta_h\zeta_m-\gamma_h\eta_m)^2+(\gamma_h\xi_m-\alpha_h\zeta_m)^2+(\alpha_h\eta_m-\beta_h\xi_m)^2]^{\frac{1}{2}}=\sin 2(\vartheta_m-\vartheta_B).$$

Substituting (11.7) one obtains, with $\alpha_h=\cos 2\vartheta_B$,

$$\vartheta_m-\vartheta_B=\tfrac{1}{2}(1-\alpha_h)\phi=\sin^2\vartheta_B.\phi. \qquad \ldots\ldots(11.8)$$

This is *Faxén's formula*, mentioned above; it gives only the "longitudinal" deviation of the extra spot from the Bragg reflection while (11.7) gives both components of the deviation.

Born and Sarginson (1941) have worked out the intensity formula (8.7) under the same simplifying assumption. The extra spots should have an elliptic shape, the minor axis being inclined by an angle $\tan^{-1}(\gamma/\beta)$ to the line connecting the Bragg and extra spots. But these formulae are hardly of any value, since the anisotropy of the scattering power around the reciprocal lattice points is of deciding importance, as will be seen immediately. On the other hand, the total intensity of an extra spot will, to some extent, be independent of the anisotropy. Therefore the formula for the three cubic Bravais lattices will be given here; one has, for the simple and the face-centred lattice,

$$I = I_0 \sigma_0 N |F|^2 \frac{16\pi\hbar^2 \sin^2\vartheta_B}{mk\Theta a^2} \frac{T}{\Theta} e^{-T/T_k} \log \frac{\sinh(\Theta/T)}{\sinh(w_0\Theta/T)}, \qquad \ldots\ldots(11.9)$$

where

$$w_0 = \frac{\pi a}{\sqrt{3}\lambda}(\gamma_k \phi). \qquad \ldots\ldots(11.10)$$

ϕ means again the angular deviation of the crystal from the Bragg position (direction cosines α_k, β_k, γ_k; for spots near to the axis of rotation $\gamma_k = \sin 2\vartheta_B$ approximately). For the body-centred lattice the expression for I has to be divided by 3, that for w_0 multiplied by $\sqrt{3}$.

The formula (11.9) has been applied by Mrs. Lonsdale (see her *Report*) to explain the observations by Laval of the dependence of the intensity on the Miller indices of the spot considered. The Bragg angle ϑ_B is expressed in terms of h_1, h_2, h_3; $|F|^2$ is a function of ϑ_B, and a "Lorentz" factor has to be added accounting for the finite angular width of the incident beam.

The validity of the assumption of isotropic scattering power has already been doubted by Waller in his Dissertation (1925); when he derives Faxén's formula he remarks that "the approximation is rough". In fact, even for elastically isotropic crystals, where $\epsilon = 0$, (9.6) reduces, not to some function of q^2 but to

$$d(\mathbf{q}) = \frac{4\pi\rho}{q^4}\left(\frac{(\mathbf{h}\,.\,\mathbf{q})^2}{c_{11}} + \frac{(\mathbf{h}\wedge\mathbf{q})^2}{c_{44}}\right), \qquad \ldots\ldots(11.11)$$

where $\mathbf{h}$ and $\mathbf{q}$ are vectors with the components h_1, h_2, h_3, and q_1, q_2, q_3, respectively.

If ψ is the angle between $\mathbf{h}$ and $\mathbf{q}$, then $d(\mathbf{q}) = \text{const.}$ leads to

$$q^2 = \text{const.}\left(\frac{\cos^2\psi}{c_{11}} + \frac{\sin^2\psi}{c_{44}}\right); \qquad \ldots\ldots(11.12)$$

these are oval surfaces produced by rotating the curve (11.12) in a plane through the vector $\mathbf{h}$ around this vector. They reduce to spheres only if $c_{11} = c_{44}$; but this, together with $\epsilon = c_{11} - c_{12} - 2c_{44} = 0$, would mean

$$c_{11} = -c_{12} = c_{44},$$

hence either the compressibility $\frac{1}{3}(c_{11} + 2c_{12})$ or the shearing coefficient c_{44} would be negative, which is impossible (Jahn and Lonsdale, 1942).

Therefore it is necessary to use (11.11) even for elastically isotropic crystals; but if (11.5) are introduced into (11.11) the result is extremely complicated and can be treated only by approximate methods.

Zachariasen (1940, 41) assumes—as mentioned above—the Cauchy relation $c_{12}=c_{44}$; then the anisotropy parameter (9.3) is

$$\epsilon = c_{11} - 3c_{12}.$$

He considers two special cases, namely (i) $c_{11}-3c_{12}=0$, or $\epsilon=0$, and (ii) $c_{11}\gg 3c_{12}$. In case (i) one has the simple formula (11.12) with $c_{11}=3c_{12}$, $c_{11}=c_{12}$, hence

$$d(\mathbf{q}) = \frac{4\pi\rho}{3c_{12}a^2q^2}(\cos^2\psi + 3\sin^2\psi) = \frac{4\pi\rho}{3c_{12}a^2}\,\frac{1+2\cos^2\psi'}{q^2}, \quad \ldots\ldots (11.13)$$

where $\psi' = \frac{\pi}{2} - \psi$.

The cross-section of this intensity distribution cut out by the plane of incidence can be expressed as an explicit function of ψ' with the help of figure 3, which

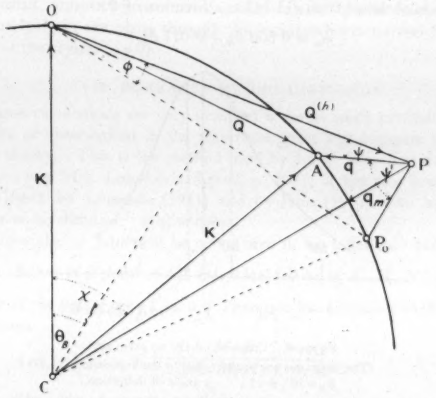


Figure 3.

ϑ_B Bragg's glancing angle of reflection.
χ total angle of deflection.
ϕ angle of rotation of crystal.

$$|\mathbf{K}| = |\mathbf{K}'| = 2\pi/\lambda.$$

$$\mathrm{OPP_0} = \psi + \psi' = \pi/2.$$

$$\mathrm{PP_0} = -\mathrm{OP_0}\,.\,\phi = -\phi\,.\,2\,\frac{2\pi}{\lambda}\sin\vartheta_B \text{ (small)}.$$

Minimum of q $\quad q_m = \mathrm{PP_0}\cos\vartheta_B = -\frac{2\pi}{\lambda}\sin 2\vartheta_B\,.\,\phi,$

and $\quad q_m = q\cos(\psi' - \vartheta_B).$

Hence $\quad \frac{1}{\psi^2} = \left(\frac{\lambda}{2\pi}\right)^2 \frac{\cos^2(\psi'-\vartheta_B)}{\sin^2 2\vartheta_B\,\phi^2}.$

Further: $\quad 2\vartheta_B - \chi = \frac{\lambda}{2\pi}\mathrm{AP_0} = \frac{\lambda}{2\pi}(\mathrm{AP}\sin(\psi'-\vartheta_B) + \mathrm{PP_0}\sin\vartheta_B)$

$$= -\phi(\sin 2\vartheta_B\,.\tan(\psi'-\vartheta_B) + 2\sin^2\vartheta_B),$$

$$\chi - 2\vartheta_B = \frac{2\phi}{1+\cot\psi'\cot\vartheta_B}.$$

shows the sphere of propagation, centre C, and a reciprocal lattice point for two positions of the crystal: P_0 for the Bragg position of selective reflection; P for a displaced position (the symbols are explained and some simple relations given under the figure). Thus $d(\mathfrak{q})$ goes over into a function of ψ':

$$d(\psi') = \frac{4\pi\rho}{3c_{12}}\left(\frac{\lambda}{2\pi a}\right)^2 \frac{\cos^2(\psi' - \vartheta_B)(1+2\cos^2\psi')}{\phi^2 \sin^2 2\vartheta_B}, \quad \ldots\ldots(11.14)$$

and as

$$\chi = 2\vartheta_B + \frac{2\phi}{1 + \cot\psi' \cot\vartheta_B}, \quad \ldots\ldots(11.15)$$

one can also consider it as a function of the total angle of deflection χ (or of the deviation $\chi - 2\vartheta_B$ from the Bragg reflection). This curve $d(\chi)$ is plotted in figure 4 for $\vartheta_B = 30°$ and $\phi = 1°$. The value ψ'_m of ψ for which $d(\psi')$ has a maximum is obtained from (11.14) as a function of ϑ:

$$\psi'_m = 0{\cdot}600\,\vartheta_B + 0{\cdot}017\,\vartheta_B^2 + \ldots\ldots \quad \ldots\ldots(11.16)$$

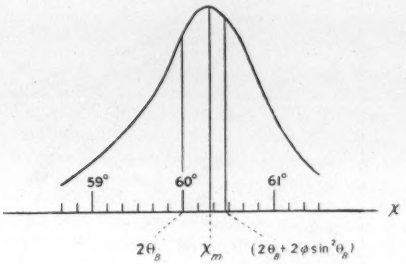


Figure 4. Cross-section of an extra spot.
(The ordinates are proportional to the expression (11.17).)
$\vartheta_B = 30°$, $\phi = 1°$; χ angle of deflection.
The ordinate marked ϑ_B corresponds to the Bragg reflection, χ_m to the diffuse maximum $2\vartheta_B + 2\phi \sin^2\vartheta_B$ to the maximum calculated for isotropic scattering power $q^2 = \text{min}$.

Using only the first term in (11.16) the position of the maximum is given by $\chi_m = 2\vartheta_m$, where

$$\vartheta_m = \vartheta_B + \frac{\phi}{1 + \cot(0{\cdot}6\,\vartheta_B)\cot\vartheta_B} \quad \ldots\ldots(11.17)$$

This must be compared with (11.8), which can be written in the form

$$\vartheta_m = \vartheta_B + \frac{\phi}{1 + \cot^2\vartheta_B} \quad \ldots\ldots(11.18)$$

This angle is a little larger than the correct one (11.17). Both are marked in the figure together with the position $2\vartheta_B$ of the Bragg reflection.

Comparison of this theory with experiment is discussed in Mrs. Lonsdale's ***Report.***

Zachariasen's results in the case (ii), $c_{11} \gg 3c_{12}$ are more complicated. In the case of reflection from a cube face (indices $(h, 0, 0)$) he obtains

$$d(\psi') \propto \frac{\cos^2(\vartheta_B - \psi')}{\phi^3 \sin^2 2\vartheta_B \left[1 + \frac{c_{11}-c_{12}}{c_{12}} \sin^2 \psi'\right]}$$

In the general case he finds three peaks of intensity which he connects with observations on diamond made by Raman and Nilakantan (1940) (see Mrs. Lonsdale's *Report*).

Miss Sarginson (1942) has developed formulae for the position of the maxima of diffuse scattering in terms of the variables ξ, η, ζ, used above in (11.5), assuming that $\epsilon = 0$ and $c_{12} + c_{44} \ll c_{11}$. She has also treated the case of finite, though small, anisotropy parameter ϵ, for certain special cases, namely, for reflections lying in the plane through the incident beam normal to the axis of rotation of the crystal ($\eta = 0$).

§ 12. SURFACES OF ISO-DIFFUSION

As these calculations are very involved it seems more profitable to express the results of observations in the reciprocal space and compare them directly with the theory. This is the method used by Laval, who first discovered the extra spots (see Mrs. Lonsdale's *Report*, p. 269); it has also been applied by Weigle (1941), by Lonsdale (1941) and by Jahn (1942), who has calculated "surfaces of iso-diffusion" $d(\mathbf{q}) = \text{const.}$

The formulae of Jahn will be given here in his notation. He denotes the direction cosines of a point in the reciprocal lattice by L, M, $N\left(L = \frac{h_1}{h}, \ldots\right)$ and those of the vector $\mathbf{q}$ by l, m, n. Then one has according to (9.6) (omitting the constant)

$$q^2 = \frac{A(L, M, N;\ l, m, n)}{A_0(L, M, N;\ l, m, n)}, \qquad \ldots\ldots (12.1)$$

with

$$\left.\begin{aligned} A &= L^2\{c_{44}[c_{44}l^2 + c_{11}(m^2+n^2)] + \epsilon(c_{11}+c_{12})m^2n^2\} + \ldots \\ &\quad - 2MN(c_{12}+c_{44})mn(c_{44}+\epsilon l^2) - \ldots \\ A_0 &= c_{11}c_{44}^2 + \epsilon c_{44}(c_{11}+c_{12})^2(m^2n^2 + n^2l^2 + l^2m^2) \\ &\quad + \epsilon^2(c_{11}+2c_{12}+c_{44})l^2m^2n^2 \end{aligned}\right\} \qquad \ldots\ldots (12.2)$$

(See Mrs. Lonsdale's *Report*, formula (9), p. 285.)

Jahn gives special values of these formulae for simple cases in which the influence of the anisotropy is put in evidence, and he applies them to the strongly anisotropic crystal of sodium. The elastic constants are (Quimby and Siegel, 1938):

$$c_{12} = 0{\cdot}760\, c_{11}, \quad c_{44} = 0{\cdot}914\, c_{11} \quad (\text{at } 200^\circ \text{ K.}),$$

hence

$$\epsilon = -1{\cdot}588\, c_{11}.$$

Some of the shapes obtained are shown in figure 5. In a second publication Jahn (1942 b) gives detailed tables and curves in stereographic projection,

$$l = \frac{2R}{1+R^2}\cos\phi, \quad m = \frac{2R}{1+R^2}\sin\phi, \quad n = \frac{1-R^2}{1+R^2}$$

Figure 6 shows some of his curves, which represent lines of equal q^2 on the surface of iso-diffusion, projected on the unit sphere surrounding a specified reciprocal lattice point. The strong anisotropy is obvious, and the results are

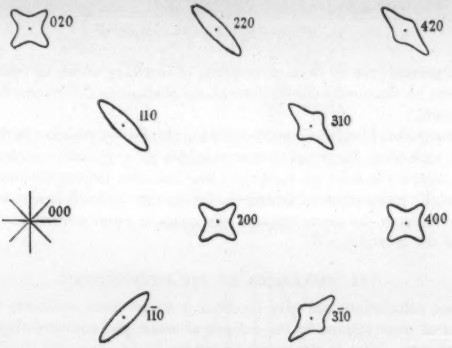


Figure 5.

(Reproduced, by kind permission, from the 'Proceedings of the Royal Society'.)

in excellent agreement with observations made by Laval (1939) and by Lonsdale and Smith (1942), which can be considered as decisive in favour of the thermal theory (see the detailed discussion in Mrs. Lonsdale's *Report*, p. 286).

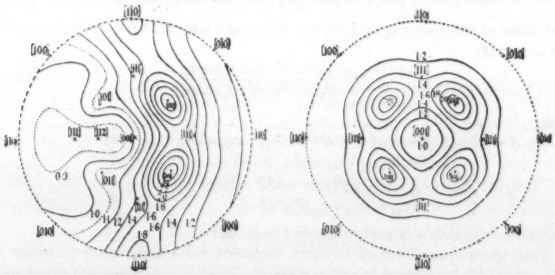


Figure 6.

(Reproduced, by kind permission, from the 'Proceedings of the Royal Society'.)

Weigle (1941) calculates the surfaces of iso-diffusion (which he calls "poles de diffusion") for diamond and illustrates his results by figures which clearly show the anisotropy of these surfaces.

§ 13. EXTENSIONS OF THE THEORY

The theory, as developed by Jahn, is valid only in the vicinity of the reciprocal lattice points. But the intermediate regions are also interesting. Experiments

show the existence of streamers connecting different Laue spots. From the theoretical standpoint weak intensity maxima of scattering power midway between the reciprocal lattice points can be expected; for in these directions the optical branches of the vibrational spectrum have minima of frequency (as is well known from the linear model of a crystal), hence the function $\phi(\omega^2) \sim kT/\omega^2$ has maxima. This region can be investigated either by using lattice models, for which $\mathscr{D}(\mathbf{Q})$ is explicitly known, e.g., for the face-centred cubic lattice ($6.23\,h$), or by numerical methods. For some crystals the elements of the dynamical matrix $\mathscr{D}(\mathbf{Q})$ are tabulated for a set of points $\mathbf{Q}$ in the reciprocal

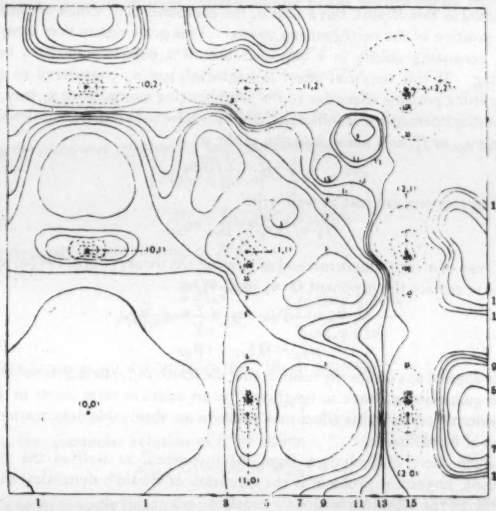


Figure 7. Lines of equi-diffusion in the plane $Q_z=0$ of the reciprocal space (Q_x, Q_y, Q_z). The figures attached to the lines are the values of $d(Q)/kT$ if the elastic constants are made dimensionless by dividing them by $e^2/a^4=2{\cdot}40\times10^{11}$ dynes/cm.^{2} (e = charge of the electron, a = lattice constant).

(*Reproduced, by kind permission, from 'Nature'.*)

space; for each of these points the elements of $\mathscr{P}=\phi(\mathscr{D}(\mathbf{Q}))$ can be found. Such calculations of $\mathscr{D}(\mathbf{Q})$ are made by Iona (1941) for KCl treated as a simple cubic lattice, by neglecting the mass difference between K and Cl, and by Kellermann (1940) for NaCl, where the different masses are taken into account and the whole six-dimensional matrix $\mathscr{D}(\mathbf{Q})$ is given for a great number of points. Both these methods have been applied to KCl by Born and Begbie (1943). The results differ only in detail. They are illustrated by figure 7, which represents the cross-section through the reciprocal space, calculated from the theoretical formula ($6.23\,h$) for $\mathscr{D}(\mathbf{Q})$. The lines in the diagram are the intersections of the surface of iso-diffusion with the plane $Q_z=0$. In the neighbourhood of the

lattice points the curves of Jahn can be recognized; but it is seen that there is also a definite structure of scattering power in the regions between the lattice points, in the shape of ridges with flat secondary maxima, connecting the steep peaks in the lattice points. Corresponding "streamers" have been observed (see Mrs. Lonsdale's *Report*, p. 272); but a quantitative comparison of the theory with experiment must be postponed until detailed intensity measurements are available.

In §2 (2.3) it was indicated that the Raman effect proper may influence the scattering. The polarizability of the single atoms is in general not a scalar, as assumed in this *Report*, but a tensor, the components of which are influenced by the position of the neighbouring atoms. This gives rise to very complicated effects, consisting chiefly in a depolarization of a polarized incident beam by scattering. If this tensorial effect is neglected and α_k considered as a scalar there is still a possible effect due to the neighbouring atoms, since α_k may depend on their displacements $\mathbf{u}$ as indicated in (2.3). The simplest assumption consists in taking α_k, or f_k^0, as a linear function of the $\mathbf{u}$:

$$f_k^0 \to f_k^0 + \sum_{k'} f_{kk'}^{(1)} \mathbf{u}_{k'}; \qquad \ldots\ldots(13.1)$$

instead of this one can just as well write

$$f_k^0 \to f_k^0 \exp\left[\sum_{k'} \boldsymbol{\beta}_{kk'} \cdot \mathbf{u}_{k'}\right]. \qquad \ldots\ldots(13.2)$$

This allows of a very simple method of treating the theory (Born, 1942a) in which one has to replace the exponent $\mathbf{Q} \cdot \mathbf{u}_k$ in (2.4) by

$$\mathbf{Q} \cdot \mathbf{u}_k - i \sum_{k'} \boldsymbol{\beta}_{kk'} \cdot \mathbf{u}_{k'} = \sum_{k'} \mu_{kk'} \cdot \mathbf{u}_{k'},$$

where

$$\mu_{kk'} = \mathbf{Q}\delta_{kk'} - i\boldsymbol{\beta}_{kk'}. \qquad \ldots\ldots(13.3)$$

If this is applied to a lattice the result is that the scattering matrix $\mathscr{S}$ is not changed, but the quantities $\mathbf{w}_k$ have to be altered by an additive term linear in $\boldsymbol{\beta}_{kk'}(\mathbf{Q})$. In polyatomic crystals this effect may produce an observable deformation of the surfaces of iso-diffusion.

A much more difficult but, from the theoretical as well as the practical standpoint, important problem is the extension of Ewald's dynamical theory of scattering to the vibrating lattice.

APPENDIX I

Proof of the formula $(e^{i\omega t\xi})_{\mathrm{av}} = e^{-k\omega^2 t^2/\omega^2}$

(A) *Classical statistics.* The energy function of a linear harmonic oscillator of unit mass is

$$\epsilon = \tfrac{1}{2}p^2 + \tfrac{1}{2}\omega^2\xi^2. \qquad \ldots\ldots(\text{A.I } 1)$$

Hence, with $\bar{\epsilon} = kT$,

$$(e^{i\omega t\xi})_{\mathrm{av}} = \frac{\iint e^{i\omega t\xi} e^{-\epsilon/kT}\, d\xi\, dp}{\iint e^{-\epsilon/kT}\, d\xi\, dp} = \frac{\int_{-\infty}^{+\infty} e^{i\omega t - \frac{\omega^2}{2k}\xi^2}\, d\xi}{\int_{-\infty}^{+\infty} e^{-\frac{\omega^2}{2k}\xi^2}\, d\xi};$$

hence by elementary integration,

$$(e^{i\omega t\xi})_{\mathrm{av}} = e^{-\omega^2 t^2/2\omega^2} = e^{-k\omega^2 t^2/2\omega^2}. \qquad \ldots\ldots(\text{A.I } 2)$$

(B) *Quantum statistics.* Let the Hamiltonian of any mechanical system be denoted by $\hbar\omega\mathscr{H}(p, q)$, where $\hbar\omega$ is a constant energy and $\mathscr{H}$ dimensionless. Let $\psi_1(q)$, $\psi_2(q)$, . . . be a complete orthonormal system of functions. Following Dirac, the function

$$\rho(q, q'/\beta) = \sum_n \psi_n^*(q') e^{-\beta\mathscr{H}} \psi_n(q) \qquad \text{.....(A.13)}$$

$\left(\text{where } \beta = \dfrac{\hbar\omega}{kT}\right)$ is called the "density matrix"; it determines the statistical (thermodynamical) properties of the system. If, for instance, ψ_1, ψ_2, . . . are the eigenfunctions of $\mathscr{H}$ itself, corresponding to the energy values E_1, E_2, . . ., then

$$\mathscr{H}\psi_n = E_n\psi_n, \quad e^{-\beta\mathscr{H}}\psi_n = e^{-\beta E_n}\psi_n;$$

hence
$$\rho(q, q'/\beta) = \sum_n \psi_n^*(q')\psi_n(q) e^{-\beta E_n}$$

and
$$\int \rho(q, q/\beta) dq = \sum_n e^{-\beta E_n}.$$

In the same way one has for any function $F(q)$

$$\int F(q)\rho(q, q/\beta)\, dq = \sum_n F_{nn} e^{-\beta E_n},$$

where
$$F_{nn} = \int \psi_n^*(q) F(q) \psi_n(q)\, dq$$

is the diagonal matrix element of $F(q)$.

Hence

$$F(q)_{\text{av}} = \frac{\sum_n F_{nn} e^{-\beta E_n}}{\sum_n e^{-\beta E_n}} = \frac{\int F(q)\rho(q, q/\beta)\, dq}{\int \rho(q, q/\beta)\, dq} \qquad \text{.....(A.14)}$$

The function $\rho(q, q'\beta)$ satisfies the "Bloch equation"

$$\frac{\partial\rho}{\partial\beta} = -\mathscr{H}\rho \qquad \text{.....(A.15)}$$

and is that particular solution of it for which

$$\rho(q, q'/0) = \sum_n \psi_n^*(q')\psi_n(q) = \delta(q - q'). \qquad \text{.....(A.16)}$$

In order to apply this to the oscillator, we substitute in (A.11) for p the operator $-i\hbar\dfrac{\partial}{\partial\xi}$ and write

$$\mathscr{H} = \frac{\epsilon}{\omega\hbar} = \frac{1}{\omega\hbar}\left(-\hbar^2\frac{\partial^2}{\partial\xi^2} + \omega^2\xi^2\right) = -\frac{\partial^2}{\partial q^2} + q^2, \qquad \text{.....(A.17)}$$

where
$$q = \xi(\omega/\hbar)^{\frac{1}{2}} \qquad \text{.....(A.18)}$$

Then the equation (A.15) becomes

$$\frac{\partial\rho}{\partial\beta} = \frac{\partial^2\rho}{\partial q^2} - q^2\rho,$$

and the solution satisfying (A.16) is

$$\rho(q, q'/\beta) = \frac{1}{(2\pi\sinh\beta)^{\frac{1}{2}}} e^{-\frac{1}{4}\{(q+q')^2\tanh\frac{1}{2}\beta + (q-q')^2\coth\frac{1}{2}\beta\}},$$

as can be directly verified by differentiation.

Hence, with $\bar{\epsilon} = \frac{1}{2}\hbar\omega \coth \frac{1}{2}\beta$ (see (4.8)),

$$\rho(q, q|\beta) = \frac{e^{-q^2 \tanh \frac{1}{2}\beta}}{(2\pi \sinh \beta)^{\frac{1}{2}}} = \frac{e^{-\frac{\omega}{2\bar{\epsilon}} q^2}}{(2\pi \sinh \beta)^{\frac{1}{2}}}. \qquad \text{.....(A.I 9)}$$

The average (A.I 4) is therefore the same as in the classical case:

$$(e^{i\omega t})_{\mathrm{Av}} = e^{-\frac{1}{2}\omega^2 q_0^2/\omega^2} = e^{-\frac{1}{2}(\omega^2 f_0/\omega)^2 \coth \frac{1}{2}\beta} \qquad \text{.....(A.I 10)}$$

APPENDIX II

Proof of the formula $\sum_j e^j f(\omega_j^2)\tilde{e}^j = f(\mathcal{D})$

Let $\mathbf{e}^j$ $(j = 1, 2, \ldots N)$ be a set of orthonormal vectors (components e_{kx}^j, $k = 1, 2, \ldots N$; x represents x, y, z) in $3N$-dimensional space. Each vector $\mathbf{e}^j$ can be represented as one-column matrix e^j or as one-row matrix $\tilde{e}^j$.

Then the condition of ortho-normality can be written in either of the two forms

$$\tilde{e}^j e^{j'} = \delta_{jj'}, \qquad \sum_j e^j \tilde{e}^j = 1, \qquad \text{.....(A.II 1)}$$

where 1 is the unit matrix in $3N$ dimensions.

Let $\lambda_1, \lambda_2, \ldots \lambda_{3N}$ be a set of real numbers, then for any real n

$$\Lambda^{(n)} = \sum_j \lambda_j^n e^j \tilde{e}^j \qquad \text{.....(A.II 2)}$$

is a symmetric matrix (elements $\Lambda^{(n)}_{kx,\,k'y} = \sum \lambda_j^n e_{kx}^j e_{k'y}^j$), and one has with (A.II 1)

$$\Lambda^{(n)}\Lambda = \sum_{jj'} \lambda_j^n \lambda_{j'} e^j \tilde{e}^j e^{j'} \tilde{e}^{j'} = \sum_j \lambda_j^{n+1} e^j \tilde{e}^j = \Lambda^{(n+1)},$$

where Λ is written instead of $\Lambda^{(1)}$. It follows that

$$\Lambda^{(n)} = \Lambda^n, \quad n = 0, \pm 1, \pm 2, \ldots \qquad \text{.....(A.II 3)}$$

Now let $\lambda_j = \omega_j^2$, e^j be the eigenvalues and eigenvectors of the dynamical matrix $\mathcal{D}$ $(= \tilde{\mathcal{D}})$, so that

$$\mathcal{D} e^j = \lambda_j e^j. \qquad \text{.....(A.II 4)}$$

Then by multiplying from the right with $\tilde{e}^j$ and summing over j, using (A.II 1),

$$\mathcal{D} = \sum_j \lambda_j e^j \tilde{e}^j = \Lambda.$$

Hence from (A.II 3)

$$\mathcal{D}^n = \Lambda^n = \Lambda^{(n)} = \sum_j \lambda_j^n e^j \tilde{e}^j = \sum_j \omega_j^{2n} e^j \tilde{e}^j. \qquad \text{.....(A.II 5)}$$

If, now, a function $f(x)$ can be expanded into a power series

$$f(x) = \sum_n a_n x^n \qquad \text{.....(A.II 6)}$$

of positive or negative powers, then one has

$$f(\mathcal{D}) = \sum_n a_n \mathcal{D}^n = \sum_j a_n \sum_j \omega_j^{2n} e^j \tilde{e}^j$$

$$= \sum_j \Big(\sum_n a_n \omega_j^{2n}\Big) e^j \tilde{e}^j,$$

or

$$f(\mathcal{D}) = \sum_j f(\omega_j^2) e^j \tilde{e}^j. \qquad \text{.....(A.II 7)}$$

APPENDIX III

Notation and symbols

ϑ_B = Bragg's glancing angle of selective reflection (monochromatic).

ϑ_L = Laue's glancing angle of selective reflection (white radiation).

χ = angle of deflection of scattered beam relative to the forward direction of primary beam.

χ_m = angle of deflection for diffuse maximum (extra spot).

λ = wave-length, ω = frequency (per 2π sec.) } (taken as being unaltered by scattering).

$\mathbf{s}$ = unit vector in the direction, $\mathbf{K} = \frac{2\pi}{\lambda}\mathbf{s}$ = wave vector } of the incident beam.

$\mathbf{s}'$ =, $\mathbf{K}' = \frac{2\pi}{\lambda}\mathbf{s}'$ = } the same for the scattered beam.

$\mathbf{Q} = \mathbf{K}' - \mathbf{K}$ = wave vector of an elastic wave ($\mathbf{Q}/2\pi$ = position vector in the reciprocal lattice).

$\mathfrak{E} = \mathfrak{E}_0 e^{i\omega t}$ = electric field of incident beam at centre of scattering system.

M^0_{rs} = matrix element of electric moment M^0 between states r, s.

m_k = mass, $\mathbf{r}_k$ = position vector, $\mathbf{u}_k$ = displacement vector, $\mathbf{v}_k = \mathbf{u}_k m_k^{\frac{1}{2}}$ = reduced displacement, $\alpha_k = \alpha_0 f_k^0$ = polarizability } of atom k.

α_0 = polarizability of free electron.

f_k^0 = scattering factor of atom k at rest ($T = 0$).

$f_k = f_k^0 e^{-M_k}$ = scattering factor of atom k modified by temperature.

M_k = Debye exponent for atom k.

M = Debye exponent for crystal [intensity of scattering proportional to $\exp(-2M)$].

I_0 = incident intensity.

I = scattered intensity.

$\sigma = I/I_0$ = scattering power.

σ_0 = scattering power of a free electron.

n = number of atoms { per molecule. / per cell of a crystal.

N = number of cells participating in scattering.

ρ = density.

a = lattice constant, c_{11}, c_{12}, c_{44} = elastic constants } of a cubic crystal.

ω_j = eigenfrequency, $\mathbf{e}_k^j$ = eigenvector of atom k, ξ_j = normal co-ordinate } for the normal mode j.

$\mathfrak{D} = \begin{cases} (\mathfrak{D}^{kk'}_{xy}) = \text{dynamical matrix of a molecule.} \\ (\mathfrak{D}^{l-l'}_{xy}) = \text{dynamical matrix of a crystal.} \end{cases}$

$\mathfrak{S} = \phi(\mathfrak{D})$ = scattering matrix.

$$\phi = \frac{\bar{\epsilon}}{\omega^2} = \frac{\hbar}{\omega}\left(\frac{1}{2} + \frac{1}{e^{\hbar\omega/kT} - 1}\right) = \frac{\hbar}{2\omega}\coth\left(\frac{\hbar\omega}{2kT}\right)$$

$\bar{\epsilon}$ = Planck's function for the average energy of an oscillator.

$\mathfrak{D}(\mathbf{Q})$, $\mathfrak{S}(\mathbf{Q})$ = irreducible parts of the matrices $\mathfrak{D}$, $\mathfrak{S}$ belonging to the elastic wave $\mathbf{Q}$.

$\left.\begin{matrix} c_D = \text{mean velocity of sound} \\ \Theta = \text{characteristic temperature} \end{matrix}\right\}$ according to Debye.

T_h = inversion temperature for the selective reflection h_1, h_2, h_3.

α, β, γ = components of $\mathbf{s}'$, when those of $\mathbf{s}$ are $1, 0, 0$.

Φ = position angle of crystal.

$\left.\begin{matrix} \alpha_h, \beta_h, \gamma_h \\ \Phi_h \end{matrix}\right\}$ = the same for the selective reflection h_1, h_2, h_3.

$\xi = \alpha - \alpha_h, \eta = \beta - \beta_h, \zeta = \gamma - \gamma_h$ = deviation of the direction cosines.

$\phi = \Phi - \Phi_h$ = deviation of the position angle.

REFERENCES

BLACKMAN, M., 1935 a. *Proc. Roy. Soc.* A, **148,** 365 and 384.
BLACKMAN, M., 1935 b. *Proc. Roy. Soc.* A, **149,** 117 and 126.
BLACKMAN, M., 1937 a. *Proc. Roy. Soc.* A, **159,** 416.
BLACKMAN, M., 1937 b. *Proc. Camb. Phil. Soc.* **33,** 94.
BLACKMAN, M., 1941. *Rep. Prog. Phys.* **8,** 11.
BLACKMAN, M., 1942. *Proc. Phys. Soc.* **54,** 377.
BORN, M., 1923. *Atomtheorie des festen Zustandes* (Leipzig: Teubner).
BORN, M., 1940. *Proc. Camb. Phil. Soc.* **36,** 160.
BORN, M., 1941. *Nature, Lond.*, **147,** 674.
BORN, M., 1942 a. *Proc. Roy. Soc.* A, **180,** 397.
BORN, M., 1942 b. *Proc. Phys. Soc.* **54,** 362.
BORN, M. and BEGBIE, G. H., 1943. *Nature, Lond.*, **152,** 19.
BORN, M. and BLACKMAN, M., 1942. *Nature, Lond.*, **150,** 55.
BORN, M. and GOEPPERT-MAYER, M., 1933. *Handb. d. Phys.* **24,** ii (2nd ed.), chap. 4. p. 623 (Berlin: Springer).
BORN, M. and v. KÁRMÁN, TH., 1912. *Phys. Z.* **13,** 297.
BORN, M. and v. KÁRMÁN, TH., 1913. *Phys. Z.* **14,** 15, 65.
BORN, M. and LONSDALE, K., 1942. *Nature, Lond.*, **150,** 490.
BORN, M., LONSDALE, K., PRESTON, G. D. and SMITH, H., 1942. *Nature, Lond.*, **149,** 373.
BORN, M. and MISRA, R. D., 1940. *Proc. Camb. Phil. Soc.* **36,** 173.
BORN, M. and SARGINSON, K., 1941. *Proc. Roy. Soc.* A, **179,** 69.
BRAGG, Sir W. H., 1940. *Nature, Lond.*, **146,** 509.
BRAGG, Sir W. H., 1941 a. *Nature, Lond.*, **142,** 780.
BRAGG, Sir W. H., 1941 b. *Proc. Roy. Soc.* A, **179,** 51.
BRAGG, Sir W. H., 1941 c. *Proc. Roy. Soc.* A, **179,** 94.
BRILLOUIN, L., 1922. *Ann. Phys., Paris,* **17,** 88.
DARWIN, C. G., 1914. *Phil. Mag.* **27,** 315.
DEBYE, P., 1913. *Verh. dtsch. phys. Ges.* **15,** 678, 738, 857.
DEBYE, P., 1914. *Ann. Phys., Lpz.*, **43,** 49.
DEBYE, P., 1927. *Phys. Z.* **28,** 135.
EWALD, P. P., 1913. *Phys. Z.* **14,** 465.
EWALD, P. P., 1933. *Handb. d. Phys.* xxiii (2nd ed.), chap. 4. pp. 279, 312.

FAXÉN, H., 1923. *Z. Phys.* **17**, 266.
IONA, M., 1942. *Phys. Rev.* **60**, 822.
JAHN, H. A., 1941. *Nature, Lond.*, **147**, 511.
JAHN, H. A., 1942 a. *Proc. Roy. Soc.* A, **178**, 320.
JAHN, H. A., 1942 b. *Proc. Roy. Soc.* A, **180**, 476.
JAHN, H. A. and LONSDALE, K., 1942. *Phys. Rev.* **61**, 375.
JAUNCEY, G. E. M., 1939. *Phys. Rev.* **56**, 647.
JAUNCEY, G. E. M., 1941. *Nature, Lond.*, **147**, 146.
JAUNCEY, G. E. M. and BALTZER, O. J., 1940. *Phys. Rev.* **58**, 1116.
JAUNCEY, G. E. M. and HARVEY, G. G., 1940. *Phys. Rev.* **58**, 179.
KELLERMANN, E. W., 1940. *Philos. Trans. Roy. Soc.* A, **238**, 513.
KERN, J., 1914. *Phys. Z.* **15**, 337.
LAUE, M. v., 1913. *Ann. Phys., Lpz.*, **42**, 1561.
LAUE, M. v., 1926. *Ann. Phys., Lpz.*, **81**, 877.
LAUE, M. v., 1941. *Röntgenstrahl-Interferenzen* (Leipzig: Akad. Verlags Ges.).
LAVAL, J., 1938. *C.R. Acad. Sci., Paris*, **207**, 169.
LAVAL, J., 1939 a. *C.R. Acad. Sci., Paris*, **208**, 1512.
LAVAL, J., 1939 b. *Bull. Soc. franç. Miner.* **62**, 137.
LEDERMANN, W., 1943. *Nature, Lond.*, **151**, 197. A detailed account is in press—*Proc. Roy. Soc.* A.
LONSDALE, K., 1942. *Proc. Phys. Soc.* **54**, 314.
LONSDALE, K. and SMITH, H., 1941. *Proc. Roy. Soc.* A, **179**, 8.
MAUGUIN, C. and LAVAL, J., 1939. *C.R. Acad. Sci., Paris*, **208**, 1446, 1512.
OTT, H., 1935. *Ann. Phys., Lpz.*, **23**, 169.
PENG, H. W. and POWER, S. C., 1942. *Proc. Camb. Phil. Soc.* **38**, 67.
PLACZEK, G., 1934. *Handb. d. Radiologie*, vi, **2**, 209.
PRESTON, G. D., 1938. *Nature, Lond.*, **142**, 570.
PRESTON, G. D., 1939 a. *Nature, Lond.*, **143**, 76.
PRESTON, G. D., 1939 b. *Proc. Roy. Soc.* A, **172**, 116.
PRESTON, G. D., 1940. *Proc. Phys. Soc.* **52**, 77.
PRESTON, G. D., 1941 a. *Nature, Lond.*, **147**, 358 and 467.
PRESTON, G. D., 1941 b. *Proc. Roy. Soc.* A, **179**, 1.
QUIMBY, S. L. and SIEGEL, S., 1938. *Phys. Rev.* **54**, 293.
RAMAN, Sir C. V., 1941 a. *Proc. Indian Acad. Sci., Bangalore*, **13**, 1.
RAMAN, Sir C. V., 1941 b. *Curr. Sci.* **10**, 515; **11**, 473.
RAMAN, Sir C. V., 1942. *Proc. Indian Acad. Sci., Bangalore*, **14**, 65.
RAMAN, Sir C. V. and NAGENDRA NATH, N. S., 1940. *Proc. Indian Acad. Sci., Bangalore*, **12**, 83, 427.
RAMAN, Sir C. V. and NILAKANTAN, P., 1940 a. *Nature, Lond.*, **145**, 667, 860; **146**, 523, 686.
RAMAN, Sir C. V. and NILAKANTAN, P., 1940 b. *Curr. Sci.* **9**, 165.
RAMAN, Sir C. V. and NILAKANTAN, P., 1940 c. *Proc. Indian Acad. Sci., Bangalore*, **11**, 379, 389 and 398; **12**, 141.
RAMAN, Sir C. V. and NILAKANTAN, P., 1941. *Curr. Sci.* **10**, 241.
SARGINSON, K., 1942. *Proc. Roy. Soc.* A, **180**, 305.
SCHRÖDINGER, E., 1914 a. *Phys. Z.* **15**, 79.
SCHRÖDINGER, E., 1914 b. *Phys. Z.* **15**, 497.
WALLER, I., 1923. *Z. Phys.* **17**, 398.
WALLER, I., 1925. *Diss.* (Uppsala).
WALLER, I., 1928. *Z. Phys.* **51**, 213.
WEIGLE, J., 1941. *Helv. phys. Acta*, **14**, 595.
WEIGLE, J., 1942 a. *Helv. phys. Acta*, **15**, 162.
WEIGLE, J., 1942 b. *C.R. Soc. Phys. Hist. nat. Genève*, **59**, 56.
WEIGLE, J. and BLEULER, K., 1942. *Helv. phys. Acta*, **15**, 445 and 553.
WEIGLE, J. and SMITH, C. S., 1942. *Phys. Rev.* **61**, 23.
ZACHARIASEN, W. H., 1940. *Phys. Rev.* **57**, 597, 795.
ZACHARIASEN, W. H., 1941. *Phys. Rev.* **59**, 766, 860, 909.

THE PHYSICS OF PAINTINGS

By F. IAN G. RAWLINS,
The National Gallery, London

§1. INTRODUCTION

In the great days of Italian and Flemish painting profound interest and concern were taken in artists' materials and the methods by which pictures were composed and constructed. The former amounted, in effect, to a rather primitive study of what is now called properties of matter, and the latter is broadly known as technique. As years went by there set in a gradual decline in mastery over the substances used in the studio, with consequent deterioration in lasting qualities and general stability. An example is provided by the habit of Sir Joshua Reynolds in using bitumen. By this means he sought to obtain a certain mellowness commonly associated with the great masters of the past; but, in fact, he greatly reduced the permanence of his pictures, and, generally speaking, made many of them next to impossible to repair. This is due to the presence of this bituminous layer, which is never really solid (except when it is exceedingly brittle), and which is the cause of continuous flaking and blistering. Compared with this, the great painters of Florence, and their successors in other Italian Schools, as well as the van Eycks in Holland, expended infinite pains in experimenting with and perfecting their pigments, binding media and primings (Laurie, 1937). Many of the Italian masters worked in tempera, an aqueous medium in which the binding was provided largely by yolk of egg. Cenino Cenini discusses at length in his book, published about 1400, the exact details of this technique, and his receipts have never been surpassed. In the Low Countries a different approach was gradually being developed, the use of oil-painting. Not infrequently the so-called discovery of oil-painting is ascribed to the van Eycks (Maroger, 1932). No doubt they were supreme masters, but there is little evidence that a new method appeared ready-made at their hands; rather a gradual transition set in, and it seems as if both tempera and oil were used together, either as a mixture of emulsions, or possibly in alternate layers. By the year 1600, approximately, tempera had practically ceased to be employed, and oil media became general. In modern times a renewed interest in tempera has arisen, and more recently still attempts have been made to use some of the new long-chain compounds—the polyvinyls—as media. Thus the physics of artists' materials, empirical and unconscious though it be, has shown great vitality and considerable powers of evolution. In following sections of this *Report*, other aspects of a picture's construction, viewed as an object rather than as a work of art, will be discussed in detail. It has been found helpful to separate individual properties of materials, like paints, resins, and varnishes, from bulk characteristics of the painting as a whole. There is a close connection between the two, and the task of the physicist, amongst other things, is to examine the relations which exist between them. In so doing he will help largely in the task of conservation and in providing data for the close

study of technique. These aspects are also of consequence in deciding upon the best physical environment for pictures, a matter in which scientific aid is being sought more fully than in the past.

It is well to recollect too that some of the most important paintings in the world are not "pictures" at all. Commonly, "pictures" are thought of as objects which are more or less portable. These are known professionally as "easel paintings". But another class comprise "mural paintings", in which are found examples of the superb frescoes which adorned the churches and monasteries of the middle ages, not to mention buildings of classical antiquity. The methods of the physicist will differ, of course, according to which of these two main categories he is considering. But one feature will be common to all his work. He is supposedly dealing with priceless and irreplaceable things, and, therefore, "safety first" will be his constant guide. Translated into modern language, this means, in effect, non-destructive testing. It is here that physical methods and appliances score heavily over chemical ones. By means of special radiations and optical appliances a great deal can be learnt without touching a painting in any way. In the last resort, a minute fragment may be detached for spectrographic or micro-chemical determination if all other means fail.

But, broadly speaking, experience has shown that a knowledge of composition of material—such as chemistry can give—is of less consequence both to scholars and to curators than information about construction, technique and condition such as physics can more readily produce. It is probably this feeling which has led to the establishment, in recent years, of several physical laboratories attached to picture galleries, or in association with them.

We have considered thus far, and very briefly, how it is that the physicist can be of service in the advance of knowledge about paintings themselves. There is, however, another opportunity open, as already indicated, and that concerns the provision and maintenance of a proper environment in museums and galleries. In our great cities the presence of dust and fog are hazards, and, in general, the question of air-conditioning in some form or another needs close attention. Ways and means by which this might be accomplished will be discussed. In practice, however, one cannot do this wholly satisfactorily without giving due weight to the needs of, and general amenity for, visitors, and this aspect will therefore be borne in mind. The same applies to lighting, both natural and artificial. Here the scientist comes into contact with designers and architects, as well as with the fundamental ideas of directors and curators. Finally, the bibliography of the subject is somewhat larger than is commonly supposed. In time of war it is naturally restricted, but before 1939 it was growing rapidly, and may be expected to do so again when happier times return. In this *Report* §§ 2, 3 and 4 are more strictly laboratory matters, whereas §§ 5 and 6 deal with the results of tests carried out on a laboratory scale perhaps, but afterwards translated into public practice.

§ 2. PICTURE STRUCTURE AND MECHANICS

It is a perfectly natural reaction to look for some time at a great picture without realizing in the least how, as a physical object, it is made. When, however, we begin to examine its construction we are confronted with a system of

considerable complexity. Details have differed throughout the ages, and in some respects modern methods are simpler than the ancient ones, but, broadly, there is one principle underlying them all. This is, that a painting (whether "easel" or "mural") is a *stratified object*, consisting essentially of a series of layers (Rawlins, 1937a). For the scientist this is somewhat fortunate, for he is accustomed to deal with structures of this kind in geology, petrology, crystallography and, indeed, in several branches of physical chemistry. In addition to a certain mental familiarity, he will even design his apparatus from the start with this fundamental idea in view.

Thus a painting is built up in general as follows (starting from the back). It comprises:

(i) the Support,
(ii) the Ground,
(iii) the Paint film,
(iv) the Varnish film.

The next section will deal with these in detail: the present purpose is to consider them more approximately, especially in their relation to the whole assemblage. The advantage in doing this is that it makes their practical importance clearer than would be the case if their minute characteristics were described first.

The *support* is the substance (wood, canvas, etc., for "easel" paintings, or stone for frescoes) upon which the whole work is based. The *ground*, sometimes called the preparation, is a layer, the purpose of which is to provide as perfect a surface as possible upon which to paint. It is usually a calcium salt, spread with great skill and evenness over the support. The *paint film* is the design itself—in fact, the "picture" It is a complex phase of pigment plus binding medium (the latter is sometimes termed "vehicle"). Finally, the *varnish film* is the coating applied for protective purposes when the design is thoroughly dry. These four expressions (support, ground, paint film, varnish film) were agreed upon internationally before the war, and may therefore be accepted as a nomenclature in which there is general acquiescence.

Two important points must be mentioned. The first is that it is not necessary for all four strata actually to be present. In a water-colour painting, for example, there is usually no ground and no varnish film. The support is paper, and directly upon it the paint film is applied. On the other hand, some of the layers may be complex: for instance, it is not unusual to find up to eight or nine separate applications of ground, the finest, naturally, being at the top. The second point is that, in any event, there is a real physical and chemical distinction between the support and the other layers above it, whereas within the latter themselves the difference is more a matter of purpose and order of application than it is of kind. In a word, ground, paint film and varnish film are all essentially paint. They differ a great deal in species, but not in nature.

On this conception of the stratified structure of all paintings has arisen the subject of picture-mechanics. This is a comparatively new outlook, which has proved fruitful in various ways. It is concerned with the stresses and strains which paintings suffer both in the normal course of their existence and on

account of exceptional events like transit, unsuitable storage, and so forth. Experiments have been undertaken to test the results of fatigue: in practice this meant subjecting a series of old trial panels to alternating forces and conditions, brought about both mechanically and by violent fluctuations in relative humidity (Anon., 1932). Several effects were obtained by these measures, resembling well-known troubles ordinarily the result of the passage of time. Flaking and warping naturally set in, but perhaps the most significant outcome was that the layers above the support together combined to inhibit to some extent the changes which the latter would naturally undergo: in other words, a considerable proportion of the strains were taken up by the ground, paint film and, perhaps, even by the varnish film. Their resistance is by no means small, but once it is overcome, deterioration of a serious kind may set in (the various layers coming away individually from each other and from the support, either separately or together), and the panel or canvas will approach a condition almost beyond repair. This type of breakdown had been known for many years before these experiments induced it artificially. The present writer has obtained some indirect evidence of a similar kind from a qualitative study of the materials concerned, based upon rheological principles of flow and deformation (Rawlins, 1941a). The physical properties of artists' materials, e.g. size, oil, paint, varnish and so on, were considered from this point of view, with the result that viscosity and compression in particular were found to be larger in general for the layers above the support than the corresponding values for wood and canvas. The suggestion is that these strata do, in fact, provide a large contribution to the reaction to its surroundings of the whole picture, when once equilibrium is disturbed, or even in the course of construction. It will appear in §5 of this *Report* how important it is to avoid relative humidities corresponding to the plastic state. Such conditions may be excellent if fabrics and textiles are being "worked" (as in factory practice), but not for long-period storage or exhibition. A problem somewhat akin is an endeavour to obtain thermal equilibrium between a picture and the wall upon which it may be hung (Beckett, 1937). All galleries are apt to contain some cold spots, and experiments undertaken some years ago under the direction of the Department of Scientific and Industrial Research demonstrated clearly what improvement could be effected by fixing metal foil upon the back of a painting under such conditions.

The fine system of fissures to be seen upon paintings of classical art—and even upon some more modern work—is characteristic, and may possess a certain beauty. The effect is now usually called "crackle" (van Dantzig, 1936). There are several kinds, which can be distinguished by their scale, depth, and shape of cross-section. Extreme cases are those which arise from the shrinkage of the support away from the ground, and the very superficial, but irritating, kind due to varnish too hastily applied. Patient microscopical examination is frequently able to unravel the various species, and even makes it possible to report whether they are genuine, or have been created purposely as part of a forgery.

In several instances it may be stated with confidence that the crackle is pure, and a true criterion of age. Certain media show characteristic patterns almost impossible to reproduce, though, of course, attempts have been made to do so.

An entirely different type of crackle arises from "volcanic" action on a minute scale. Certain pigments liberate small quantities of gas in the process of time. These exudations may sometimes be violent enough to form a cañon of microscopic magnitude, fundamentally like that occasioned by a great natural geological upheaval. All these happenings are studied under the general subject of picture mechanics: they are of interest both on historical grounds and as aids in conservation.

§3. THE PHYSICS OF STRUCTURAL DETAIL

(i) *The support*

The support is the substance upon which the whole painting is based. For "easel" pictures, wood, canvas, paper, metal, and recently fabricated boards have been used. An example of the compound layer already mentioned is the Italian habit, common before the year 1400, of applying cloth over the wood panel. A good soaking in glue made from goat or sheep parchment provided excellent adhesion, and the practice was altogether a sound one. The Renaissance brought canvas more into fashion—pictures were often very large, and the reduction in weight was an advantage. The physics of wood has been progressively studied. The chief property of moment is its strongly anisotropic behaviour with respect to moisture-content. Forsaith (1926) has found that, in general, a loss of 1% by weight in air-dried wood produced a radial shrinkage of approximately 0·2%, and a tangential shrinkage of about 0·3%. The axial, or longitudinal, change in dimensions is practically nil. The importance of these facts is in relation to the atmospheric condition to be aimed at for pictures: it will be discussed in detail in §5. Meanwhile, however, it may be noted that there are many galleries where the relative humidity fluctuations correspond to a moisture-change in the wood of about 8%: this involves an expansion or contraction in the extreme case of some 2% in length—in fact a very considerable strain. Canvas behaves in the opposite sense, but, for practical purposes, the magnitude of the effect is not very different. Vellum becomes embrittled at a temperature of about 80° F., which sets a limit to drying by the application of heat (Anon., 1939).

(ii) *The ground*

The purpose of this substratum has already been mentioned. Technically, it is often known as *gesso*, the Italian for gypsum. There are other types of ground, but broadly they are aqueous white inerts, including chalk or zinc oxide. In Italian panels the ground is commonly glue mixed with calcined plaster of Paris. There are at least two varieties, a coarse thick *gesso grosso*, and a fine (almost microcrystalline) *gesso sottile*. Normally, the former is found in the lower layer, and the latter above. For the application of gold leaf, a species of clay has been used since mediaeval times or earlier. This is known as "bole". There is a white bole, much the same as kaolin, and a red bole, an earthy mineral akin to ochre. It will appear in a later paragraph that the ground is not only of significance for our knowledge of picture construction: for the radiologist it is essential that he should be aware of the different types in existence, or he may be lulled into very erroneous conclusions in studying his x-ray photographs.

In the United States (Gettens, 1937), experiments of great delicacy have been performed which show in microscope sections the stratified character of a painting, and reveal many layers of gesso of progressive degrees of aggregation. It is also clear from these observations how the particles of pigment tend to lock into the ground, giving mechanical stability to the paint film above.

(iii) *The paint film (and pigments)*

Physically, paint is an aggregate of pigment particles, dispersed in a film-forming substance (medium or vehicle) which anchors them firmly in position. Evidently the properties of such a material will depend upon particle size, relative indices of refraction of pigment and medium and a number of other factors. For example, the well-known effect in tempera whereby the colours become lighter on drying is due to the evaporation of water and its replacement by air in the medium (egg-yolk).

Structurally, for the special purposes of museum work, paint is divided into the following three classes, depending upon the relative proportions of pigment and medium:

(a) *Granular paint*, in which the pigment content is high.

(b) *Vehicular paint*, in which the quantities of pigment and medium are both moderate. (This class includes much of common occurrence in all techniques.)

(c) *Pellicular paint*, in which medium predominates over pigment. A limiting case of this category is formed by the glazes. In these, the pigment content may be vanishingly small; they consist almost wholly of oil, with possibly a coloured dye suspended in it.

At low magnifications (about × 10) with an adequate binocular microscope it is usually possible to distinguish between these three divisions (and certain sub-divisions within them) quite easily. The advantage of the scheme, which is due to G. L. Stout (1938), is that it is readily applicable to what the curato and restorer usually want to know, and further, is independent of the chemical composition of the phases.

(iv) *The varnish film*

Its main purpose is purely protective, against mechanical damage, moisture, dust and smoke. Pastel paintings are sometimes sprayed with a fixative, a natural resin, to hold the pigment particles firmly in position. The common varnishes are dammar, sandarac and copal, together with the new polyvinyl compounds. These latter have the advantage of great transparency and whiteness, but their permeability to moisture is high. Formerly varnish films were applied by brush almost universally; now they are often sprayed. A technique has been developed by H. Ruhemann * whereby a wax finish is applied over a varnish layer below, another example of a compound structure. The resistance to atmospheric deterioration is very good, and the surface may even be carefully

* See Stout and Cross (1937), and also the Proceedings of Conference reviewed in *Les Dossiers de l'Office International des Musées*, 2, *Documents sur la Conservation des Printures* (Paris": Institute International de Co-operation Intellectuelle, 1933), pp. 12, 13 and 35.

washed with soap and water if necessary. It should be noticed that no varnish films are thick enough to act as perceptible thermal insulators; in so far as humidity is concerned, some are fairly efficient barriers (Gettens and Bigelow, 1932), offering considerable resistance to moisture fluctuations. The removal of discoloured varnish layers is a preliminary to more extensive restoration, and may, in fact, be all that is needed in some cases, without more drastic treatment. Naturally, the restorer's problem is to obtain solvent action without damage to the paint film below. Stout and Cross (1937) list a number of compounds in increasing order of severity—paraffin hydrocarbons are the mildest, glycol ethers and ketones the most violent.

(v) *True fresco and "secco"*

Some of the greatest masterpieces in existence belong to the category of fresco painting. The manipulative skill in working with true fresco must have been prodigious, since speed is the essence of the process when it comes to executing the actual design. A very brief description is included here for convenience and for the sake of completeness. The support is normally stone and the ground is plaster, a compound layer consisting of at least two subsidiary parts, coarse and fine. In true fresco the medium is lime water, or baryta water in fresco secco. Sometimes both lime and baryta are found in the latter. Carbonization of the alkaline substances takes place during fixation. In both types of fresco the ground is wet and caustic. In Great Britain, true fresco is scarcely possible, for climatic reasons, but in Italy much of it has stood up to conditions excellently.

The conditions underlying fresco technique are fundamentally those of the phase rule, and it is noteworthy to what an extent they must have been empirically appreciated by the great exponents of the art. Some modern directions, aimed at recapturing the essentials of the process in the light of present-day knowledge, are given in a recent circular issued in the United States (Anon., 1940).

This concludes a brief study of the strata composing a painting. With this in mind, it will be easier to appreciate the bulk reactions to which reference has already been made in discussing picture mechanics, and also the interpretation of results by means of the special radiations, described in the next section. Beyond that, a knowledge of the characteristics of these layers is essential for the choice and control of environment.

§4. THE GALLERY LABORATORY

(i) *Aims and limitations*

In one sense, laboratories in connection with paintings have existed for centuries. Mediaeval prints and drawings show workshops and studios replete with retorts, bellows, ovens, grinding devices and much of the rest of the paraphernalia of the craftsman. Only in recent years, however, have physical methods come to be directly applied to gallery work (Plenderleith, 1941 a, 41 b). The task is now seen to comprise a number of inter-related but distinct investigations. Thus, the physics of artists' materials outlined in previous sections of

this *Report* are always making progress as technological research advances. The gallery laboratory, however, as now understood, is mainly concerned with paintings themselves as they exist to-day, sometimes supremely precious, unique and irreplaceable, and, at the least, interesting and important objects of study. From these considerations two aims become reasonably clear. One is the application of physics to problems of conservation, and the other is to the extension of our knowledge of the technique of the great masters of classical art. An example of the former is provided by an enquiry into the condition of a picture, and whether it is capable of being cleaned, restored, or, in general, brought back to health, if it is a serious "invalid".

Preliminary steps in such an enquiry will be to examine the surface, edges and back minutely with a hand lens (magnification about × 6) to see to what extent the main strata are in proper contact with each other, or have pulled away as the result of mechanical, physical or chemical damage. The presence of mildew may also be noticed. Photographs, including, perhaps, life-size details, will be taken of any specially important or interesting areas, including cracks, flakes, blisters, "bloomed" varnish and so forth. A visual examination will then be made in ultra-violet light (ultra-violet photographs can, of course, be taken if it is thought necessary to obtain permanent records). This will probably reveal a great deal about the varnish, but very little about the rest of the composition. The reason is that most varnishes fluoresce strongly in the ultra-violet, masking any effect from the paint film. Naturally, if the varnish is later removed, this limitation no longer applies, and the pigments are open to identification by the usual chemical reactions. However, the fluorescence from the varnish tells the scientist whether it is intact all over the surface, and whether, as occasionally happens, any repaint has been added over it. The restorer (Lyon, 1934a) will sometimes proceed to remove the varnish under ultra-violet control, so that he may know exactly and sharply by how much, and precisely where, his solvents are acting. Furthermore, if he wishes to be certain that a particular part is really free from varnish, this ultra-violet exploration is very useful, and fairly rapid.

It is quite likely, however, that for museum reasons it is undesirable to remove the varnish at all, at least in the early stages of diagnosis. In this case infra-red photography will probably give a valuable indication of the condition of the paint film (Lyon, 1934b), due to the high transmissivity of varnish, even if markedly discoloured, for this part of the spectrum. The reflectivities of most pigments for these frequencies are now known, sufficiently at least for broad recognition (Farnsworth, 1939). Next will follow x-ray analysis (Wilde, 1931; Rawlins, 1939a). In favourable instances the deep structures of the picture will be laid bare, and at times important, and even dramatic, discoveries have been made. These, however, should not be counted upon in advance, as a number of adverse factors, either singly or together, may go far to defeat the experiment. Such factors are heavy lead grounds, excessive thickness of wood (including cradling), too even a density of pigments, or a paint film of an exceptionally thin character. In any event, however, the radiologist will be confronted with a record showing several strata superposed upon each other, and the work of interpreting it correctly is largely a matter of patience and

experience. It is, here perhaps, more than anywhere else, that the physicist must possess a sufficient knowledge of pictures as such, historically as well as technically. Otherwise he is severely handicapped.

The radiation tests accomplished, microscopical observations of the "crackle" and condition of the paint film itself will follow, using polarized light to reduce flare. In addition, if restoration and cleaning are decided upon, colorimetric analysis of important areas and details may be called for, before and after the restorer's work, in order to register the change of tone, and to retain records for the gallery files in case of future criticism or comment.

One major point must always be borne in mind. The decision to restore rests ultimately upon aesthetic values—that is, whether the picture will or will not be more beautiful after treatment than before (Clark, 1937–39). The scientist's part is limited to objective statements as to what is there and what is not. In a similar way, he can give no judgement upon stylistic questions, though the laboratory has often contributed appreciably to their solution (Rawlins, 1939 b). Plate 1 illustrates the way in which physical methods are being used in the second category already mentioned, i.e., in the study of the techniques of the old masters. The infra-red photograph (plate 1 (*a*)) is of the hands of *Saint Francis* from the great *Demidoff Altarpiece* painted by Crivelli in 1476. An ordinary panchromatic photograph (plate 1 (*b*)) is seen in comparison. Certain artists—Tura, Cossa and Crivelli (there may be others, but these three have been definitely established)—were accustomed to stress their undermodelling of limbs exceptionally—a characteristic little seen in the original, but brought to light by the penetration of the outer layers of paint by the infra-red rays. The effect, though far from obvious visually, is a most significant part of their rendering of anatomical detail. The skill demanded by this method of emphasizing the third dimension must have been altogether out of the ordinary.

The use of x rays in the gallery laboratory is illustrated by plates 2 and 3. Plates 2 (*a*) and 3 (*a*) are ordinary photographs of the heads of the two angels which form the right and left wings of the famous *Virgin of the Rocks* by Leonardo da Vinci (1452–1519); x-ray photographs of the same subjects are shown in plates 2 (*b*) and 3 (*b*). These angels were for long thought to be both by Predis. It can now be appreciated how different they are. No. 1661 is seen by x rays to have neither drawing nor modelling, and to have nothing in common with its companion. When the original version of Leonardo's great picture was taken to France, it is likely that Predis's other angel went with it, and No. 1661 is a substitute, the work of an inferior pupil. This is one among several examples of the power of scientific research to help curators and historians. The point is that, although No. 1661 has for some time been suspect on stylistic (and other) grounds, x rays have placed the matter beyond doubt.

(ii) *Apparatus*

A general outline of laboratory activities in a picture gallery is now before the reader. The next step is to indicate in more detail how they are carried out, or, in other words, to describe the necessary apparatus (Laurie, 1932; Ritchie, 1935; Rawlins, 1935). With the case of the "good" and "bad" angels fresh in mind, it is natural to begin with x-ray plant. The prime essentials are simple

Plate 1 (*a*). Hands of *Saint Francis*. Detail from N.G. No. 788 by Crivelli. Infra-red photograph showing the artist's characteristic sub-surface shading.

Plate 1 (*b*). The same hands as in 1 (*a*), taken on an ordinary panchromatic plate. The veins are accentuated, but the sub-surface modelling is suppressed.

(*Reproduced by courtesy of the Trustees of the National Gallery.*)

To face page 244

Plate 2 (*a*). Detail of N.G. No. 1662 by Predis. Normal panchromatic plate.

Plate 2 (*b*). X-ray photograph of N.G. No. 1662 (15 ma. at 15 kv.).

(*Reproduced by courtesy of the Trustees of the National Gallery.*)

Plate 3 (*a*). Detail of N.G. No. 1661. Normal panchromatic plate.

Plate 3 (*b*). X-ray photograph of N.G. No. 1661 (16 ma. at 16 kv.).

(Reproduced by courtesy of the Trustees of the National Gallery.)

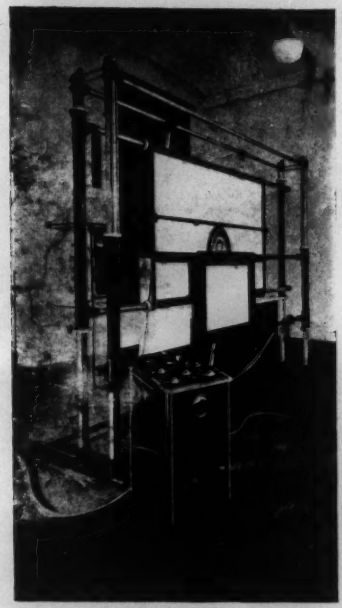


Plate 4. X-ray apparatus and easel, at the National Gallery Laboratory, with a small panel in position behind the fluorescent screen.

(*Reproduced by courtesy of the Trustees of the National Gallery and of the British Institute of Radiology.*)

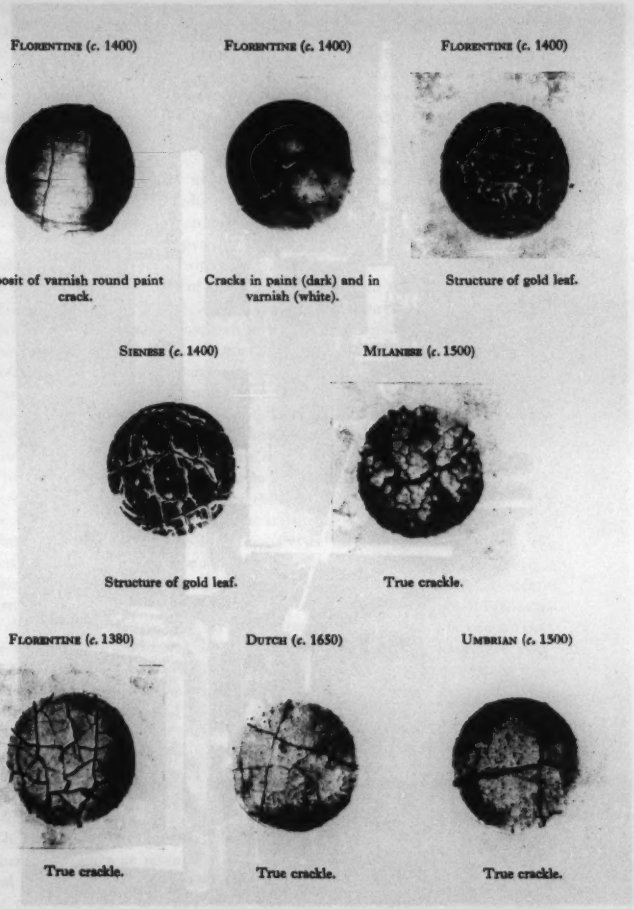


Plate 5. Photo-micrographs of paint structures. Magnification : 70 diameters.

(*Reproduced by courtesy of the Trustees of the National Gallery.*)

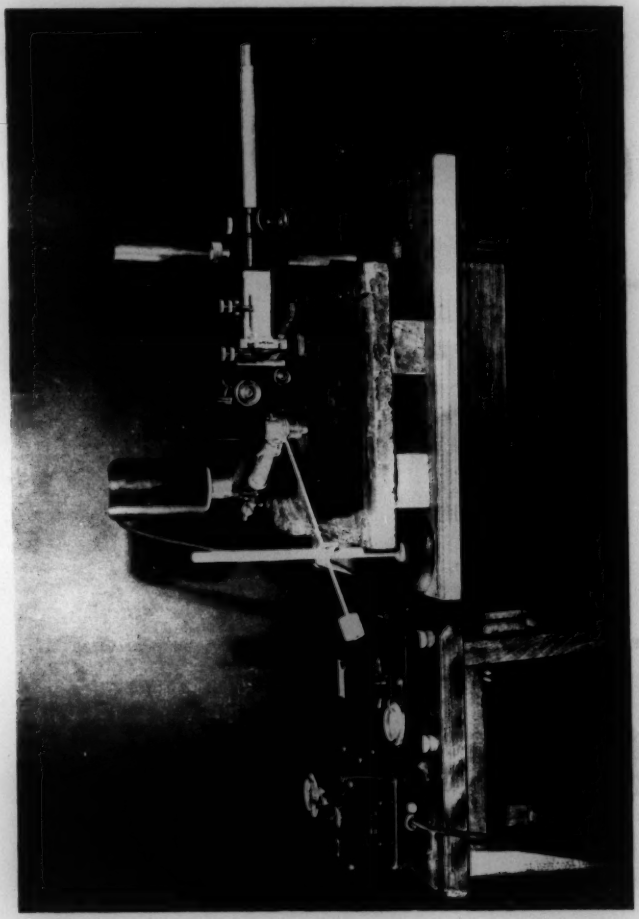


Plate 6. A small panel under examination with the travelling microscope.
(Reproduced by courtesy of the Trustees of the National Gallery and of the Royal Society of Arts.)

but important. Firstly, that of the size—and weight—of the pictures to be examined. These factors may vary (as they have in the writer's experience) from a few inches square, weighing a few ounces, to a "patient" some 10 ft. by 10 ft., scaling not far off half a ton. It has to be recollected too that it may be most desirable to investigate by screening (the usual barium tungstate screen) before taking photographs. This means that one must have all parts of a large composition readily accessible when being x-rayed. For these reasons, and many others, there is much to be said for arranging the picture vertically, as depicted in the illustration of plate 4. In addition, it is then easy to bring up other appliances, like cameras and microscopes, to it without further movement. Secondly, it is necessary to reduce the penetration of the x rays to a low limit, if fine detail and good gradation are to be achieved. In the writer's apparatus the potential is 10–15 kv. (depending upon thickness of support, nature of the pigments and ground, etc.), with a current of some 15–20 ma. A graduated switchboard provides a series of small potential steps so that conditions can be finely adjusted for each case. Since, in contradistinction to much medical practice, the subject is static, long exposures are permissible. Perhaps it should be mentioned here that calculations and observations in this country and elsewhere have agreed in neither anticipating nor finding any damage to pictures as a result of x-ray tests being applied to them. In this connection it is entertaining to remember that certain "professors" in Germany were at one time active in denouncing radiological experiments already performed upon pictures, and forecasting terrible results to come in the future. These people, however, were little else than commercial "experts", interested in the sale and application of secret preparations, supposed to be the only hope of bringing a picture back to life after it has been "ruined" by the scientists.

Ultra-violet and infra-red apparatus needs careful consideration for the particular end in view. Since a fair amount of ultra-violet work on painting is undertaken visually, it is desirable that a standard daylight lamp (of moderate intensity, so as not to produce too violent a contrast for the eye) should be available, capable of being switched on and off rapidly, for purposes of comparison, when any specially important area is under observation. A miniature ultra-violet hand-lamp, worked off the mains with a transformer, is very convenient for "running over" a picture quickly in a dark room.

Infra-red cameras should be all-metal, and provided with a parfocal lens, to facilitate visual focusing. A fixed-focus type instrument is usually sufficient—thus saving trouble with bellows—since most of the work is concerned with details, and the image may thus be of a standard size. The present writer has designed a suitable apparatus for the purpose, which has given satisfactory results (Rawlins, 1937 b).

The importance of "crackle" has already been mentioned. Plate 5 shows a number of varieties as revealed by investigation with the microscope. Some typical appearances of gold leaf are also included. The technique involved is to use polarized light—to reduce flare—combined with a system of lenses taking light from the periphery (thus acting as a self-contained condenser). The microscope and auxiliary apparatus are shown in plate 6. Actually, any arrangement of vertical, or slightly oblique, illumination will serve, so long as it enables

sufficient light-intensity to be obtained, especially for photomicrography at modest magnifications. Naturally, in order to work over the large areas of a painting systematically, it is necessary to have a microscope arranged with an exceptionally long travel. Sometimes this can be accomplished by normal methods, but occasionally the geometric-slide principle may be invoked.

In the United States another branch of microscopy has been developed to a high degree of efficiency for use upon paintings. This is the sectioning of the strata already described, and mounting in wax, or methacrylic resins (Gettens, 1940), for examination, as in biological practice. By these means it has been possible to preserve intact the foraminifera which in part composed the sedimentary substratum of the very early frescoes in Italy and elsewhere (Duell and Gettens, 1940). There is no doubt that historical research into these primitive civilizations has been materially advanced by these experiments, largely the work of R. J. Gettens at Harvard.

Spectrographic analysis has sometimes played an important part too. In the identification of pigments the method is rapid, certain and needs a minimum of material. The general technique developed for metallurgical testing can be used with advantage. When possible, the spectrograph is preferable to micro-chemical assay as being more refined, and capable of providing better permanent records in the shape of spectrograms. Some time ago Ostwald (1936) produced a scheme for the microchemical testing of binding media in paint films. His reactions are delicate, and in skilled hands have yielded valuable information concerning artists' methods and materials.

When a picture is to be cleaned, it is, as already indicated, desirable to possess some record of the tonal changes brought about by the removal of discoloured varnish. A simple way of achieving this is by the employment of some form of colorimeter or tintometer, suitably modified for the purpose (Rawlins, 1936, 37 c, 41 b, 42 a). The present writer has introduced this practice as a matter of routine, with encouraging results. It is readily possible not only to record the colour changes involved, but to deduce a number of coefficients which show their relative magnitude. Eye and pre-conception alike are apt to be most misleading if the picture as a whole is taken into account. It is perhaps well to add that tintometric tests upon trial pigments have been made in comparison with spectrophotometric studies of identical samples. The results display good consistency (Barnes, 1939; Rawlins, 1940).

§5. THE PHYSICAL ENVIRONMENT OF PAINTINGS

We now approach another aspect altogether of the part played by physics. This concerns the establishment and control of the general conditions under which paintings may best be kept. Obviously, a detailed knowledge of such objects themselves is an indispensable prerequisite for this work: in the absence of such information it is impossible to prescribe what is acceptable for their welfare, and what must be avoided. Some acquaintance with the subject of biophysics is needed as well. At the outset it should be made clear that paintings of all kinds have, in fact, in the past suffered much damage (some of it irreparable) from bad conditions: in addition, our big cities, in which many collections are normally located, provide far from ideal habitats, on account

of smoke and dirt (Rawlins, 1937 d). It is by no means true that, because many great masterpieces have survived for centuries, the dangers are not real ones. A peculiar hazard was reported many years ago from Italy: it concerned the deterioration of pictures in churches and monasteries due to the liberation of chlorine from the burning of bleached candles. To-day, the presence of sulphur acids in the atmosphere of towns and industrial districts is a risk for some pigments, and the presence of moisture in the air undoubtedly hastens the fading of water-colours, which occurs as a result of exposure, even to light in the short wave-length region of the visible spectrum. Vibration, too, has been suspected of initiating defects in old paint films.

From the point of view of air conditioning, the scientist may consider both panels and canvasses—to a first approximation—as a substrate of biological fibres (wood or textile) upon which rest a series of layers of film-forming substances. His aim will be to reduce the intake and output of moisture by this compound system to a minimum, since it is known that greater dimensional changes are brought about by this agency than by temperature fluctuations as such. Experience shows that a relative humidity of between 55% and 60% at a temperature of 60° F. is sound practice, the tolerance in the former being about ±5%. Temperature *per se* is not very important within reasonable limits. Its significance is on account of its connection with relative humidity. Fortunately, when a room or storage chamber is well filled with wood and canvas, a balancing action sets in, and the relative humidity in general becomes less sensitive to temperature changes than in the case of an empty space.

Researches have shown that, even at room temperatures, or a little higher, mildew is apt to form at relative humidities exceeding 68% (Groom and Panisset, 1933). This figure, then, represents an upper limit of safety for galleries and repositories. A fair protection against mildew is rapid air circulation, and this should always be provided, if necessary by installing fans and arranging louvers in doors and cupboards. Without it, mere raising of temperature is rarely of much avail.

The need for the protection of paintings against the risks of aerial bombardment has provided the physicist with a number of special problems (Rawlins, 1942 b). These war-time activities may prove of permanent value in times to come, for in consequence much relevant data are being accumulated. If a temporary repository is found in an ordinary surface building, it will probably be necessary to seal-off the storage room from outside atmospheric conditions, set up electric fans for local air circulation, and to open up the room only when the weather allows (Rawlins, 1942 c).

Deep-shelter storage, in underground workings, is a problem of a different magnitude entirely. Much of what will be needed is a matter for engineering rather than for physics (Fletcher, 1943; Rawlins, 1943 a, 43 b). But if the natural temperature below ground is low enough (and approximately constant), as it sometimes is, good conditions for paintings may be obtained by temperature-control alone, without refrigeration. But, in general, the crucial question is the rate at which the relative humidity in the storages will rise in the event of a failure of plant. Depending upon local circumstances, this may vary from a rise from, say, 60% to 80% relative humidity in a matter of minutes, or the

same increase may take days. In the former case the switching-in of emergency power supply is urgent; in the latter less so, though an increase of relative humidity to beyond 68% (Pease, Gettens and Stout, 1941) (the threshold value for mildew) may bring disaster in a comparatively short period.

§6. THE PUBLIC ASPECT

In normal times many, perhaps the majority, of first-rate collections are in some way open to the public, and their number is likely to increase. Consequently, the physicist is always somewhat concerned with the general surroundings of pictures in such circumstances. There are at least three ways in which air-conditioning may be achieved in galleries. One is to condition the building as a whole; another is to feed conditioned air from a central plant into individual frames and showcases containing the pictures; a third is to provide within frames or pedestals hygroscopically-balancing salts, like the hydrates of zinc sulphate, making each exhibit a unit in this respect. At the same time, whatever method is adopted, there is usually need for an air-washing and filtering plant, more particularly in large towns and industrial districts.

Many difficulties are commonly encountered due to the layout and architectural nature of the buildings, which were, of course, erected in the first instance with no thought of these refinements. Herein, too, reside most of the problems of gallery lighting (D.S.I.R., 1927). For daylight illumination, various optical systems have been proposed to overcome the irritating reflections arising from the presence of glass over the pictures. Artificial lighting may well be by fluorescent lamps, giving nearly perfect "daylight" and absence of shadows (Rawlins, 1942d).

It will now, perhaps, have become clear how much can be achieved by physics in relation to paintings. Much of this, as will have been seen, is of a decidedly specialist nature, needing methods and apparatus which are both novel and interesting.

GENERAL REFERENCES

ANONYMOUS, **1939.** *Manuel de la Conservation et de la Restauration des Peintures* (Paris: Office International des Musées; also Offices of "Mouseion").

BURROUGHS, A., **1938.** *Art Criticism from a Laboratory* (Boston: Little, Brown and Co.).

DE WILD, M., **1929.** *The Scientific Examination of Paintings* (London: G. Bell and Sons, Ltd.).

GETTENS, R. J. and STOUT, G. L., **1942.** *Painting Materials. A short Encyclopaedia* (New York: D. van Nostrand and Co., Inc.).

HERRINGHAM, C., **1899.** *The Book of the Art of Cennio Cenini* (London: G. Allen).

HOUWINK, R., **1937.** *Elasticity, Plasticity and Structure of Matter* (Cambridge: University Press).

LAURIE, A. P., **1926.** *The Painter's Methods and Materials* (London: Seeley, Service and Co.).

MORRELL, R. S., **1939.** *The Scientific Aspects of Artists' and Decorators' Materials* (Oxford: Clarendon Press).

NATIONAL GALLERY TRUSTEES, **1940.** *From the National Gallery Laboratory* (London: National Gallery Trustees).

RORIMER, J. J., **1931.** *Ultra-Violet Rays and Works of Art* (New York: Metropolitan Museum of Art).

THOMPSON, D., **1936.** *The Materials of Medieval Painting* (London: Allen and Unwin).

WOLTERS, C., **1938.** *Die Bedeutung der Gemäldedurchleuchtung mit Röntgen Strahlen für die Kunstgeschichte* (Frankfurt-am-Main: Prestel-Verlag.)

OTHER REFERENCES

[In the references which follow, abbreviation *T.S.* is used to denote *Technical Studies in the Field of the Fine Arts*, published by the Fogg Art Museum, Harvard University, Cambridge, Massachusetts, U.S.A. This quarterly, publication of which was initiated in 1932, was suspended on account of war conditions in 1942. It contains most of the original matter considered in this *Report*.]

ANONYMOUS, 1932. *Some Notes on Atmospheric Humidity in Relation to Works of Art* (London: Courtauld Institute of Art).

ANONYMOUS, 1939. *A.R.P. in Museums, Picture Galleries and Libraries* (London: British Museum Trustees).

ANONYMOUS, 1940. *Fresco Painting. W.P.A. Technical Series, Art Circular No.* 4 (Washington, D.C.: Federal Works Agency, W.P.A.).

BARNES, N. F., 1939. "A Spectrophotometric Study of Artists' Pigments"; *T.S.* **7**, 120–138.

BECKETT, H. E., 1937. "Condensation on Unglazed Pictures"; *J. Instn. Heat. and Vent. Engrs.* **5**, 134–137.

CLARK, K. M., 1937–39. "The Aesthetics of Restoration"; *Proc. Roy. Instn.* **30**, 382–397.

VAN DANTZIG, M. M., 1935. "Barstrorming op Schilderijen"; *Maanblad voor Beeldende Kunsten*, **12**, 232–249 (English translation in *T.S.* **4**, 165–167, 1936).

D.S.I.R., 1927. *The Natural Lighting of Picture Galleries. Technical Paper No.* 6 (London: H.M. Stationery Office).

DUELL, P. and GETTENS, R. J., 1940. "A Method of Painting in Classical Times"; *T.S.* **9**, 75–104.

FARNSWORTH, M., 1939. "Infra-Red Absorption of Paint Materials"; *T.S.* **7**, 88–98.

FLETCHER, P. T., 1943. "The Engineering Aspect of the Wartime Storage of Art Treasures"; *J. Instn. Heat. and Vent. Engrs.* (In press.)

FORSAITH, C. C., 1926. *The Technology of New York State Timbers. Technical Publication No.* 18, *N.Y. State College of Forestry* (New York; University of Syracuse).

GETTENS, R. J., 1937. "The Cross-Sectioning of Paint Films"; *T.S.* **5**, 18–22.

GETTENS, R. J., 1940. "The Use of Methyl Methacrylate in the Preparation of Polished Specimens of Friable Material"; *T.S.* **9**, 113–116.

GETTENS, R. J. and BIGELOW, E., 1932. "The Moisture Permeability of Protective Coatings"; *T.S.* **1**, 63–68.

GROOM, P. and PANISSET, T., 1933. "Penicillium crysoganum (Thom)"; *Ann. Appl. Biol.* **20**, 633–660.

LAURIE, A. P., 1932. "Un Laboratoire pour l'Examen des Peintures"; *Mouseion.* **17–18**, 119–122.

LAURIE, A. P., 1937. "The Refractive Index of a Solid Film of Linseed Oil: Rise in Refractive Index with Age"; *Proc. Roy. Soc.* A, **159**, 123–133.

LYON, A., 1934 a. "Ultra-Violet Rays as Aids to Restorers"; *T.S.* **2**, 153–157.

LYON, A., 1934 b. "Infra-Red Radiations Aid Examination of Paintings"; *T.S.* **2**, 203–212.

MAROGER, J., 1932. "Essai de Reconstitution de la Matière Picturale de Jean van Eyck"; *Mouseion*, **19**, 39–46.

OSTWALD, W., 1936. "Iconoscopic Studies"; *T.S.* **4**, 135–144.

PEASE, M., GETTENS, R. J. and STOUT, G. L., 1941. "The Problem of Mould Growth in Paintings"; *T.S.* **9**, 127–144.

PLENDERLEITH, H. J., 1941 a. "Application of Modern Physics in the Examination of Paintings"; *Nature, Lond.*, **147**, 165–166.

PLENDERLEITH, H. J., 1941 b. "Some Aspects of Museum Laboratory Work"; *Chemistry and Industry*, **60**, 807–812.

RAWLINS, F. I. G., 1935. "The Physical Laboratory at the National Gallery"; *Science Progress*, **30**, 236–242.

RAWLINS, F. I. G., 1936. "Studies in the Colorimetry of Paintings, Part 1"; *T.S.* **4**, 179–186.

RAWLINS, F. I. G., 1937 a. "The Physics and Chemistry of Paintings"; *J. R. Soc. Arts* **85**, 933–968 and 971–988.

RAWLINS, F. I. G., 1937 b. "A novel Infra-Red Camera for Art Gallery Work"; *Museums J.* **38**, 186–187.

RAWLINS, F. I. G., 1937 c. "Studies in the Colorimetry of Paintings, Part II"; *T.S.* **5**, 150–156.

RAWLINS, F. I. G., 1937 d. "Atmospheric Pollution, with Special Reference to the National Gallery"; *J. Instn. Heat. and Vent. Engrs.* **5**, 400–442.

RAWLINS, F. I. G., 1939 a. "X Rays in the Study of Pictures"; *Brit. J. Radiol.* **12**, 239–245.

RAWLINS, F. I. G., 1939 b. "Evidence: Its Nature and Place in the Scientific Examination of Paintings"; *T.S.* **8**, 75–81.

RAWLINS, F. I. G., 1940. "A 'Tintometric' Comparison of Artists' Pigments"; *T.S.* **9**, 3–10.

RAWLINS, F. I. G., 1941 a. "The Rheology of Painting"; *T.S.* **10**, 59–72.

RAWLINS, F. I. G., 1941 b. "Studies in the Colorimetry of Paintings, Part III"; *T.S.* **9**, 207–220.

RAWLINS, F. I. G., 1942 a. "Studies in the Colorimetry of Paintings: a Note in Conclusion"; *T.S.* **10**, 230–231.

RAWLINS, F. I. G., 1942 b. "The Control of Temperature and Humidity in Relation to Works of Art"; *Museums J.* **41**, 279–283.

RAWLINS, F. I. G., 1942 c. "Care of Works of Art in War-Time"; *Nature, Lond.*, **150**, 112–114.

RAWLINS, F. I. G., 1942 d. "The Physics of Art"; *J. Sci. Instrum.* **19**, 17–22.

RAWLINS, F. I. G., 1943 a. "Some Physical Aspects of the Storage of Works of Art"; *J. Instn. Heat. and Vent. Engrs.* (In press.)

RAWLINS, F. I. G., 1943 b. "The National Gallery in War-Time"; *Nature, Lond.*, **151**, 123–128.

RITCHIE, P. D., 1935. "The new Department and Laboratory of Scientific Research, Courtauld Institute of Art, London"; *T.S.* **3**, 217–219.

STOUT, G. L., 1938. "Classes of Simple Paint Structure"; *T.S.* **5**, 221–239.

STOUT, G. L. and CROSS, H. F., 1937. "Properties of Surface Films"; *T.S.* **5**, 241–249.

WILDE, J., 1931. "L'Examen des Tableaux à l'Institut Holzknecht de Vienne"; *Mouseion*, **16**, 18–25.